ESSAYS ON ECONOMIC DECISION MAKING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mahmut Kemal Ozbek

January 2014

ESSAYS ON ECONOMIC DECISION MAKING

Mahmut Kemal Ozbek, Ph.D.

Cornell University 2014

This dissertation consists of three chapters with the common ground of economic decision making, a subject which is at the heart of Economics.

In the first chapter, entitled "Decision Making with Rational Inattention", I and Maximilian Mihm study an axiomatic model of decision making with costly information processing. In this regard, we first provide a set of intuitive axioms for a decision maker's preferences over menus of acts from which she eventually makes a choice of an act. We then show that these preferences can be represented by a novel information acquisition model where choices of acts can be improved by using costly information. Our focus on preferences over menus allows us to uniquely identify the parameter's of the model from choice data: a utility function, a prior belief and an *information* cost function. The cost function, in particular, is compatible with the *Blackwell* order. Moreover, the model establishes an axiomatic foundation for the models of *rational inattention* which are widely applied in the literature.

In the second chapter, entitled "On Representation of Monotonic Preference Orders", I and Tapan Mitra investigate the relation between *scalar continuity* and representability of *monotone* preference orders in a sequence space. Scalar continuity is shown to be sufficient for representability of a monotone preference order and easy to verify in concrete examples. Generalizing this result, we show that a condition, which restricts the extent of scalar discontinuity of a monotone preference order, ensures representability. We also relate this condition to

the well-known *order dense property*, which is both necessary and sufficient for representability.

In the third chapter, entitled ”Rational Inattention and Choice of Optimal Information”, I study the choice problem of a *rationally inattentive* decision maker modeled according to the first chapter of this dissertation. In this work, I give a characterization for optimal *information choices* when the cost function is *linear*. In addition, the characterization result is applied on a simple buyer-seller model where the buyer is rationally inattentive to the riskiness of the seller. It is found that the optimal price should be non-monotonic in the degree of the buyer's *attentiveness*.

BIOGRAPHICAL SKETCH

Mahmut Kemal Ozbek obtained a Bachelor of Science degree in Mathematics from Bilkent University, Turkey in 2005. He then received a Master of Arts degree in Economics from Sabanci University, Turkey in 2007. That same year, he joined the Ph.D. program in the Department of Economics at Cornell University and in 2011, he was awarded a Master of Arts degree. Shortly after he earns his Doctor of Philosophy degree in August 2013, he will be joining University of St. Andrews, UK as a full time faculty in the Department of Economics and Finance.

To those beyond time...

ACKNOWLEDGEMENTS

Among the many people who contributed to my development as a scholar, I would like to express the deepest appreciation to my committee co-chairs, Tapan Mitra and Larry Blume. Tapan -with his kind nature and immense patience- always provided his time and support generously which made it possible for me to come along this far. I feel extremely fortunate to know him; a great mentor and an invaluable source of inspiration to me. Larry -with his challenging tasks and words of wisdom- influenced me considerably. I am indebted to him especially for being very supportive of me through the most important part of my Ph.D. studies.

I am also grateful to my committee members, Joerg Stoye and Aaron Bodoh-Creed, for their kind support. I benefited greatly from professional and personal discussions with them. I would like to also thank David Easley, Steve Coate, Kaushik Basu and Mukul Majumdar for their suggestions and comments on my work. I thank John Abowd and Jenny Wissink for their support and wise TA assignments.

Among the friends and colleagues, I owe a lot to John Gottula for being more than a great roommate. I feel lucky to know Maximilian Mihm and Samreen Malik for their close friendship. I thank Emre Ekinci, Aziz Simsir, Joerg Ohmsted and Ram Sewak Dubey for their support and understanding.

Last but not the least, I would like to thank my family; I am grateful to my uncle, Mustafa Sahin, my aunt, Leyla Sahin, and my cousins whose presence helped me to keep going further. Finally, I owe a lifelong gratitude to my mom who taught me economics the very first with the following simple, yet elegant words: *Intelligence is your sole resource; use it wisely.*

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

DECISION MAKING WITH RATIONAL INATTENTION

1.1 Introduction

In the age of the Internet, smart phones, and the 24 hour news cycle, an abundance of information is available at the push of a button. Yet introspection should suffice to attest that assimilating the wealth of information into actions is constrained by one's limitations on attending to information.[1] Information is money, attention is scarce, and thus people must allocate their attention to the information they deem most relevant to their decisions.[2]

The attention allocation problem has been recently introduced into economic analysis to explain some important phenomena such as the price stickiness puzzle, Mackowiak and Wiederholt (2009), ineartial reactions to shocks, Sims (1998, 2003) and business cycle dynamics, Veldkamp and Wolfers (2007). It has been also used in finance to explain home equity bias puzzle, Van Nieuwerburgh and Veldkamp (2010), to examine dynamic asset pricing theories, Peng and Xiong (2006) and to study models of mutual funds, Garcia and Vanden (2005) and decentralized trade, Golosov et al. (2012).[3]

The literature models the problem with a fully rational decision maker who needs to make decisions in an uncertain world. Yet prior to making a decision,

[1]In cognitive psychology, "*attention* [is] considered a resource to be allocated to particular mental processes [where] it is paid out on demand to facilitate selected aspects of information processing." Picton et al. (1986, p. 19).

[2]"... a wealth of information [means] a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it." Simon (1971)

[3]See Veldkamp (2011) for a detailed discussion of the literature modeling economic agents with limited attention, and see the references therein.

she can actively seek information in order to be able to make better choices. However, she might be unable to absorb all the information available due to her scarce attention. Despite its intuitive appeal and important implications, the literature incorporating the problem of attention allocation into decision making has not yet obtained a behavioral foundation. This work aims to fill this gap by understanding attention scarcity as a limitation on the ability to integrate information into state-contingent consequences. In this regard, we provide in a menu-choice setting a simple set of axioms from which we derive several models of decision making with scarce attention.

Our contribution to the literature can be summarized in two folds: (i) First, by focusing on preferences over menus, we are able to uniquely identify attention scarcity from choice behavior, and thereby clarify the decision theoretic foundations of the models using attention allocation problem. In doing this, we naturally provide a set of falsifiable conditions that one can test with the choice data to verify the validity of these models. (ii) Second, with our axiomatic results we are able to propose a novel model of decision making with costly attention which in turn broadens the scope of the existing models and hence might guide the applied research.

1.1.1 Preview of the decision problem

Our decision domain is a variation on the menu choice problem first proposed by Kreps (1979) to study preference for flexibility, and used extensively in the recent decision theory literature on subjective state spaces, Dekel et al. (2001), self-control problems, Gul and Pesendorfer (2001), contemplation and thinking costs, Ergin and Sarver (2010); Ortoleva (2012), minmax regret, Stoye (2011), and subjective learning, Lleras (2010); Dillinberger and Sadowski (2011).

There is a finite set, $\mathcal{S}$, of states of the world. A decision maker (DM) faces choice problems both before and after the realization of a state $s \in \mathcal{S}$. *Ex-ante* the DM chooses over menus of Anscombe-Aumann type acts. An act is a mapping from $\mathcal{S}$ into lotteries over a set of consequences. A menu is a finite collection of such acts. Limitations on attention manifests *ex-post*, when the DM processes information about the realized state to make a better choice from the menu. However, since the DM's knowledge about the state depends on her ability to process information, choices in different states of the world will now be random variables, and therefore not observable. Instead, we therefore focus on the *ex-ante* preferences over menus – which reflect preferences over the type of decision problems the DM would like to face *ex-post* – and ask when such preferences reveal the anticipated attention scarcity in an unmodeled choice from menus in the future. The timeline in Figure 1.1 summarizes the order of events.

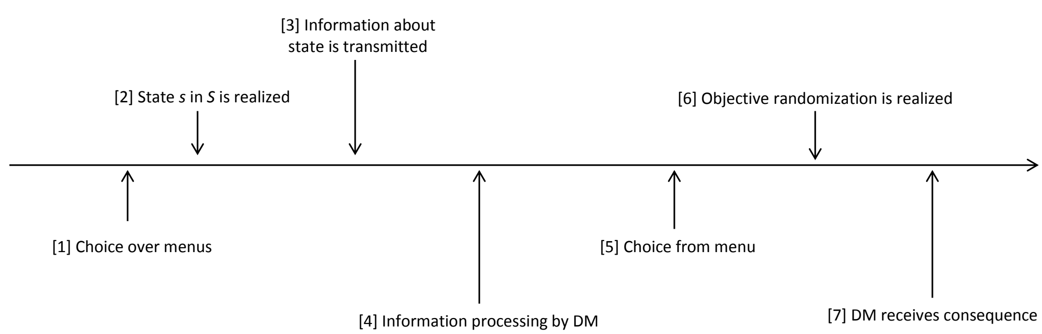

Figure 1.1: Order of decisions and resolution of uncertainty

1.1.2 Preview of the representation

For a rational agent facing a subjective cost of information processing, preferences over menus allow us to characterize the observable implications of attention scarcity. In particular, we provide an axiomatic representation result (see Theorem 1) for the preferences of a DM who evaluates a menu M according to the following utility functional:

$$V(M) = \max_{p \in \Delta(\mathcal{S} \times \Sigma)} \left\{ \int_{\mathcal{S}} \left(\int_{\Sigma} \left[\max_{f \in M} \int_{\mathcal{S}} Eu[f(s')] \, d\, p(s'|\sigma) \right] d\, p(\sigma|s) \right) d\, \pi(s) \; - \; c(p) \right\} \quad (1.1)$$

where Eu is a *von-Neumann-Morgenstern* utility function on lotteries, Σ is a countably infinite space of signals, $\Delta(\mathcal{S} \times \Sigma)$ is the set of all joint distributions on the (state $\times$ signal) product space and c is an *information cost function*.

In the background is the interpretation that the DM decides *ex-ante* on a plan for allocating scarce attention before making *ex-post* choices from menu M. Attention scarcity is captured formally by the utility functional (1.1) via the subjective cost c associated with information contained in joint distributions. The properties of c are therefore crucial to our interpretation of (1.1). In particular, there exists a unique distribution π over states such that $c(p) = \infty$ whenever $p_{\mathcal{S}} \neq \pi$, so that π admits a natural interpretation as the DM's subjective prior over the state space. The function c otherwise satisfies all basic properties demanded of a measure of information: It is (i) non-negative, convex and lower semi-continuous, (ii) equal to zero for product measures with marginal π (i.e., joint distributions which contain no posterior information about states), (iii) depends only on probabilities in the signal dimension, and (iv) satisfies the data processing inequality, a basic monotonicity condition that a measure of information "should" satisfy Cover and Thomas (2006, Chapter 2). Due to the clear parallels with modeling of information in Sims (1998, 2003), we call c an information cost function, and call the representation in (1.1) a *utility representation with variable attention*.

The signal space Σ, of course, is not a primitive in our setting; it is an endogenous construction of the representation that obtains the meaning of a signal space only through the choice of a joint distribution $p \in \Delta(\mathcal{S} \times \Sigma)$, and through the manner in which the DM utilizes the information that latent signals convey about states. In particular, the DM uses information in joint distributions as a Bayesian expected utility maximizer. Corresponding to each joint distribution $p \in \Delta(\mathcal{S} \times \Sigma)$, there is a family of likelihood functions for observations on signals, $\{p(\sigma|s)\}_{s \in \mathcal{S}}$. After observing a signal $\sigma \in \Sigma$, the DM translates the information from the likelihood of signals into a posterior over the state space, $p(s'|\sigma)$, via Bayes rule. She then chooses from menu M the act that maximizes her expected utility for the posterior. The realized state, $s \in \mathcal{S}$, affects the likelihood of signal realizations and, hence, the information which the posterior distribution contains about the true state. Since the realized state is *a priori* uncertain, the DM uses her prior π to calculate the expected value of the *ex-post* choices from the menu.

1.1.3 Preview of the axioms

The axioms that characterize utility with variable attention are surprisingly simple given the rich information theoretic content of the representation. In addition to standard weak order, monotonicity, continuity and weak independence conditions, the key behavioral postulate is a *variable attention* (VA) axiom that is related to the timing of the resolution of uncertainty (see Figure 1.1). VA is similar to the contemplation aversion axiom in Ergin and Sarver (2010), although in our domain it has a different interpretation. To motivate, we adapt an example in Ergin and Sarver (2010) to our setting.

Suppose a DM is choosing a restaurant to visit with friends. A state corresponds to the quality of dishes available at each restaurant, an act corresponds to the choice of a dish, and a menu corresponds to the choice of a restaurant. Before meeting with her friends, the DM can review restaurants online, in newspapers and magazines, to obtain information about the quality of dishes. An information processing strategy might correspond to a plan to look at certain webpages, newspaper or magazine articles, and she can actively tailor her strategy to the restaurant under consideration. For example, if her friends suggest Italian food, the DM might research "Italian dining" online, if they suggest Japanese food she might direct attention towards Japanese dining websites. But she will not be able to study all restaurant reviews; even if much of the information is publicly available, she may have insufficient cognitive resources to study them all, or simply have better things to do with her time.

Now suppose with enough effort spent studying reviews first, the DM expects to make an equally good choice at both the Italian and Japanese restaurants. She tells her friends that she is indifferent between the two restaurant options, and they decide to toss a coin. When would the DM like to know the outcome of the coin toss? If a decision over restaurants is made before she spends time on research, she can vary her attention according to the chosen restaurant. She is indifferent between the restaurants, and could reasonably be indifferent with regards to this early randomization. However, if her friends want to toss the coin later, the DM faces a more complex attention allocation problem. To make an equally well-informed choice at both restaurants she must commit more time to research, i.e. pay more attention to available information. As a result, we posit that a DM who intends to actively allocate scarce attention between multiple information sources would prefer the coin to be tossed *before* she decides how to allocate attention. In fact, the DM might strictly prefer the Italian restaurant over the delayed randomization, because late resolution only complicates her attention allocation problem. VA therefore postulates that a DM indifferent between two menus will at least weakly prefer each menu to a randomization resolved *after* variable attention has been allocated.

1.1.4 Outline of the paper

We provide additional motivation and discussion of the VA and other axioms in Section 1.2, after introducing the formal decision model. Section 1.3 then presents our main results. Theorem 1 gives the axiomatic characterization of the utility representation with variable attention in (1.1). Proposition 1 shows that *mutual information* – a measure of information that is commonly employed in applications of rational inattention – is a special case of our cost function. Theorem 2 gives a representation result for a *utility function with constrained attention*, in which inattention is modeled in terms of an *information constraint set* (a special case of an information cost function). Proposition 2 shows that the information constraint set can again be defined in terms of mutual information, so that the finite Shannon capacity model – the particular functional form suggested by Sims for rational inattention modeling – is contained as a special case.

Theorem 3 incorporates a further restriction on inattention, and characterizes a *utility function with fixed attention* that formally establishes the link between our model of decision making and models of subjective learning. Section 1.4 provides behavioral criteria for comparative statics on (in)attention. Section 1.5 discusses related literature and Section 1.6 concludes. Formal proofs are given in an Appendix.

1.2 A Decision Model with Costly Attention

This section formalizes a decision environment that is suitable for eliciting attention scarcity from choice data, and provides a number of behavioral postulates for the preferences of a rational agent for whom attending to information is costly.

1.2.1 Decision environment

Let $\mathcal{S}$ be a finite set of *states of the world* (or simply *states*), which represent all uncertainty relevant for the payoffs that the DM can receive. The power set of $\mathcal{S}$ is denoted $2^{\mathcal{S}}$. An element $E \in 2^{\mathcal{S}}$ is called an *event*.

The set $\mathcal{Z}$ denotes the set of all possible *consequences*, and $\Delta(\mathcal{Z})$ denotes the set of all simple probability distributions on $\mathcal{Z}$. An element of $\Delta(\mathcal{Z})$ is called a *lottery*. For any two lotteries $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$ and $\alpha \in [0,1]$, the lottery $\alpha\mathbb{p} + (1-\alpha)\mathbb{q}$ assigns probability $\alpha\mathbb{p}(z) + (1-\alpha)\mathbb{q}(z)$ to each consequence $z \in \mathcal{Z}$. The set of *Anscombe-Aumann acts* (or simply *acts*), denoted $\mathcal{F}$, is the set of all functions

$f : \mathcal{S} \to \Delta(Z)$.[4] We denote by $\mathcal{F}_C$ the set of all constant functions $f \in \mathcal{F}$ for which, in all $s \in \mathcal{S}$, $f(s) = \mathbb{p}$ for some $\mathbb{p} \in \Delta(\mathcal{Z})$. Mixtures on $\mathcal{F}$ are defined pointwise in the usual way: For any $f, g \in \mathcal{F}$ and $\alpha \in [0, 1]$, $\alpha f + (1-\alpha)g : s \mapsto \alpha f(s) + (1-\alpha)g(s)$.

The alternatives amongst which the DM must choose prior to the realization of a state of the world are finite subsets of $\mathcal{F}$. We denote by $\mathcal{M}$ the set of all such finite subsets of $\mathcal{F}$, and call a typical element $M \in \mathcal{M}$ a *menu*. For any $f \in \mathcal{F}$, $\{f\}$ denotes the singleton menu that contains only act f. We denote the set of all singleton menus by $\mathcal{M}_S$. With a slight abuse of notation, $\{\mathbb{p}\}$ denotes the singleton menu that contains only the constant act $\mathbb{p} \in \mathcal{F}_C$ (i.e., the constant act $s \mapsto \mathbb{p}$), and $\mathcal{M}_C$ denotes the set of all constant, singleton menus. The sets $\mathcal{M}_S$ and $\mathcal{M}_C$ are in a natural one-to-one relation with the set $\mathcal{F}$ and the set $\Delta(\mathcal{Z})$, respectively. Finally, for any $M, N \in \mathcal{M}$ and $\alpha \in [0, 1]$, $\alpha A + (1-\alpha)B \in \mathcal{M}$ denotes the menu $\{\alpha f + (1-\alpha)g \in \mathcal{F} | f \in A\ ,\ g \in B\} \in \mathcal{M}$. We observe that with respect to this mixture operation, $\mathcal{M}$ is a mixture space.

1.2.2 Axioms

We consider a DM with preferences over menus described by a binary relation $\succsim$ on $\mathcal{M}$, and interpret the statement $M \succsim N$ to mean that the DM weakly prefers menu M over menu N. Denote the asymmetric and symmetric parts of $\succsim$ by $\succ$ and $\sim$, respectively, with the usual interpretations. We now state a collection of axioms to describe preferences of a DM who is a fully rational agent facing a latent, subjective cost of information processing, but who fully anticipates limitations on information processing at the time of expressing preferences over

[4]The space of acts $\mathcal{F}$ is the reformulation by Fishburn (1970) of the decision domain first proposed by Anscombe and Aumann (1963), and commonly used in the literature on decision making under uncertainty.

menus.

Axiom 1 (Weak Order (WO)) *(i) Given $M, N \in \mathcal{M}$, then $M \succsim N$ or $N \succsim M$, (ii) Given $L, M, N \in \mathcal{M}$, then $L \succsim M$ and $M \succsim N$ implies $L \succsim N$.*

WO is the standard requirement that preferences be complete and transitive.[5] Our primary behavioral postulate on a weak preference order is directly related to the attention allocation problem of a DM. Recalling the restaurant choice example, a DM may be indifferent between two menus M and N on the basis of quite different plans about how to allocate attention before making an *ex-post* choice. It seems reasonable to posit that she would be indifferent also to an objective randomization over the menus, as long as randomness is resolved before she allocates attention. But we assume throughout that objective randomization occurs *after* choices from menus. The late resolution of objective randomization might complicate the attention allocation problem. As such, *variable attention* states that the DM will at least weakly prefer not to randomize.

Axiom 2 (Variable Attention (VA)) *Given $M, N \in \mathcal{M}$, then $M \sim N$ implies $M \succsim \alpha M + (1-\alpha)N$ for all $\alpha \in (0,1)$.*

As the name suggests, VA allows variable attention, so that attention – viewed as the cognitive resources used to process information about states of the world – can be actively allocated to the most useful information for any given choice problem. Since the attention allocated to information varies with menus, VA is permissive of preference reversals due to any randomization that

[5] Completeness in particular is a strong rationality requirement in a rich choice domain, and it could be interesting to relax the assumption to model a DM who is indecisive about the information she may be able to attend to in the future. We leave this exercise for further research.

can complicate the attention allocation problem (i.e., the DM might prefer M to N, but prefer $(1/2)N+(1/2)L$ to $(1/2)M+(1/2)L$ when mixtures of M with L complicate the attention allocation problem more than mixtures of N with L). Our next axiom restricts the scope of such reversals when randomizing with singleton menus, as with a singleton menu there is no *ex-post* choice to make. As such, randomization with singleton menus complicates the active attention allocation problem of a DM in a restricted way only.

To motivate, a singleton menu could correspond to a restaurant with a strict *table d'hôte* (i.e., no choice over dishes). Suppose now that a DM and her friends will randomize over an *à la carte* restaurant and a *table d'hôte* restaurant. The likelihood of an *à la carte* restaurant might well influence the attention allocation problem. If there is only a small probability that she ends up at the *à la carte* restaurant, it may be optimal to allocate very little attention to restaurant reviews at all. If the probability is high, there is much more incentive to try and find pertinent reviews. However, which *table d'hôte* restaurant is being considered should not affect the allocation of attention. Reviews will not improve the DM's choice at a *table d'hôte* restaurant, because the menu is fixed, and so her attention should be allocated to find information about the *à la carte* restaurant. As such, weak singleton independence postulates that for a *fixed* randomizations with a singleton menu, changing the singleton menu should not lead to preference reversals.

Axiom 3 (Weak Singleton Independence (WSI)) *Given $M,N\in\mathcal{M}$, $f,g\in\mathcal{F}$ and $\alpha\in(0,1)$, then*

$$\begin{aligned} & \alpha M+(1-\alpha)\{f\} \succsim \alpha N+(1-\alpha)\{f\} \\ \Leftrightarrow\quad & \alpha M+(1-\alpha)\{g\} \succsim \alpha N+(1-\alpha)\{g\}\ . \end{aligned} \tag{1.2}$$

To motivate our final behavioral axiom, suppose that the DM and her friends are choosing between an Italian and a Japanese restaurant, both of which have a fish and a meat option. The DM may be uncertain about whether meat or fish dishes are better in general, but if she knows that she prefers the Italian meat over the Japanese meat, and the Italian fish over the Japanese fish, then we posit that she should prefer to go to the Italian restaurant. After all, no matter what information she attends to, or what she would choose at the Japanese restaurant, the Italian restaurant has an option that she prefers. The following monotonicity condition embodies this basic principle.

Axiom 4 (Monotonicity (MON)) *Given $M, N \in \mathcal{M}$ such that for all $g \in N$ there exists $f \in M$ with $\{f(s)\} \succsim \{g(s)\}$ for all $s \in \mathcal{S}$, then $M \succsim N$.*

MON subsumes two standard dominance conditions:

(1) Restricted to singleton menus, MON reflects the usual monotonicity condition invoked in an Anscombe-Aumann settings to obtain state independent payoffs: If $\{f\}, \{g\} \in \mathcal{M}_S$, and $\{f(s)\} \in \mathcal{M}_C$ is preferred to $\{g(s)\} \in \mathcal{M}_C$ for all $s \in \mathcal{S}$, then MON implies $\{f\} \succsim \{g\}$.

(2) MON also clearly implies a "preference for flexibility" (Kreps, 1979): If $M, N \in \mathcal{M}$, and $M \supset N$, then MON implies $M \succsim N$.

Behaviorally, MON therefore assumes that the DM understands fully the state-contingent prizes associated to each act, understands the content of every menu, and is free to choose from menus. However, MON is permissive of inattention to manifest as an attribute of the DM's understanding about the occurrence of external states. In fact, we show in Section 1.4, that alternative ways of strengthening MON characterize both maximal and minimal inattention, thereby illustrating formally that MON does not restrict the inattention towards external states faced by our DM at all.

Finally, we require two more technical conditions. Both are natural adaptations of axioms in the literature, but they are not falsifiable with finite choice data.

Axiom 5 (Mixture-Continuity (MC)) *Given $f, g \in \mathcal{F}$ and $M \in \mathcal{M}$, then the sets $\{\alpha \in [0,1] : \alpha\{f\} + (1-\alpha)\{g\} \succsim M\}$ and $\{\alpha \in [0,1] : M \succsim \alpha\{f\} + (1-\alpha)\{g\}\}$ are closed.*

Axiom 6 (Unboundedness (UB)) *There exist $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$ with $\{\mathbb{p}\} \succ \{\mathbb{q}\}$ such that for all $\alpha \in (0,1)$ there exists $\mathbb{r} \in \Delta(\mathcal{Z})$ satisfying either $\{\mathbb{q}\} \succ \alpha\{\mathbb{r}\} + (1-\alpha)\{\mathbb{p}\}$ or $\alpha\{\mathbb{r}\} + (1-\alpha)\{\mathbb{q}\} \succ \{\mathbb{p}\}$.*

MC adapts the usual mixture-continuity condition employed in Anscombe-Aumann settings to menu choice. Together with WO and MON, its purpose is to guarantee the existence of constant-singleton equivalents: For all $M \in \mathcal{M}$ there exists some $\mathbb{p}_M \in \Delta(\mathcal{Z})$ such that $M \sim \{\mathbb{p}_M\}$. UB is used in a number of "arbitrage" arguments to establish uniqueness properties. It is essentially the condition used by Maccheroni et al. (2006) to identify uniquely a grounded cost function in the representation of variational preferences. We utilize their derivations in our own proofs, but rely more on UB as it also establishes the special role played by the prior π in our representation.

1.3 Representation Results

This Section states our main result: A representation theorem for preferences satisfying the axioms of Section 1.2. We also study a number of special cases of the representation in (1.1), that provide alternative restrictions on the type of (in)attention exhibited by a DM. Before stating results, we require some notation and definitions.

1.3.1 Preliminaries

Given a discrete product space $A \times B$, we denote by $\Delta(A \times B)$ the set of probability measures $p : 2^{A\times B} \to [0,1]$, and endow this space with the weak*-topology.[6] If $p \in \Delta(A \times B)$, p_A denotes the *marginal distribution* of p on A (or simply the A-marginal of p), defined by $p_A(E_A) := p(E_A \times B)$ for all $E_A \in 2^A$. Likewise, p_B denotes the B-marginal of p, defined by $p_B(E_B) := p(A \times E_B)$ for all $E_B \in 2^B$.

If $p \in \Delta(A \times B)$, $p(.|b)$ denotes the *conditional distribution* on A given $b \in B$, and defined by $p(E_A|b) := p(E_A \times b)/p_B(b)$ whenever $p_B(b) > 0$ and $p(E_A|b) = p_A(E_A)$ when $p_B(b) = 0$. Note that for $p_B(b) > 0$, $p(.|b)$ is simply the Bayesian update of the distribution p_A on observing $b \in B$, given the dependence between the random experiments A and B when they are governed by the joint probability distribution p. The condition $p(.|b) = p_A$ when $p_B(b) = 0$ is simply a convention (it is without loss of generality in the following to assume any arbitrary distribution after conditioning on the observation of an experimental outcome with marginal probability zero). Likewise, $p(.|a)$ denotes the conditional distribution

[6] In the decision theory literature the term probability measure often refers to *finitely additive* probability measures. Throughout, we follow the nomenclature in probability theory and reserve the term probability measure for measures that are also countably additive.

on B given $a \in A$ defined by $p(E_B|a) := p(a \times E_B)/p_A(a)$ whenever $p_A(a) = 0$ and $p(E_B|a) = p_B(E_B)$ when $p_A(a) = 0$.

For any joint distribution $p \in \Delta(A \times B)$, there exists a unique *product measure*, denoted $\bar{p}$, associated to p and defined by $\bar{p}(E_A \times E_B) = p_A(E_A)p_B(E_B)$ for all $E_A \in 2^A$ and $E_B \in 2^B$. We denote the set of all product measures on $A \times B$ by $\bar{\Delta}(A \times B)$ (i.e., $p \in \Delta(A \times B)$ is in $\bar{\Delta}(A \times B)$ if and only if $p(E_A \times E_B) = p_A(E_A)p_B(E_B)$ for all $(E_A, E_B) \in 2^A \times 2^B$). We note that $\bar{\Delta}(A \times B)$ is a convex subset of $\Delta(A \times B)$.

If $p, q \in \Delta(A \times B)$, p is said to be *sufficient* for q, denoted $p \succsim q$, if there is a Markov kernel[7] on $B \times 2^B$ such that $q(E_A, E_B) = \int_B \mathbb{K}(b, E_B) dp(E_A, b)$ for all $(E_A, E_B) \in 2^A \times 2^B$. If A and B are viewed as two random experiments, the Markov kernel $\mathbb{K}$ can be viewed as an *ex-post* manipulation of the observations from random experiment B. Such manipulations do not provide additional information about the random experiment A as they depend *only* on realizations of the experiment B. Blackwell (1953) therefore introduced the ordinal relation $\succsim$ on $\Delta(A \times B)$ to capture the idea that p contains at least as much information as q.[8] We denote the symmetric part of $\succsim$ by $\simeq$ (i.e., $p \simeq q$ if p and q contain equal information).

The sufficiency relation is an ordinal measure of the information that a random experiment B conveys about a random experiment A, when randomness is governed by a particular joint distribution $p \in \Delta(A \times B)$. The information theory literature has also proposed numerous cardinal *information measures*. Although any non-negative functional on $\Delta(A \times B)$ could (in principle) serve as an information measure, there are certain properties that a useful information measure

[7]See Bauer (1995, p. 305) for the definition of a Markov kernel.

[8]Although sufficiency is a purely statistical (i.e., objective) description of the information that one random experiment contains about another, Blackwell (1953) established a formal equivalence to numerous alternative definitions of informativeness motivated by decision theoretic arguments and measures of uncertainty Kihlstrom (see, e.g., 1984, for a detailed exposition).

should satisfy: (a) It should be equal to zero for product measures (under which the random experiments A and B are independent), (b) it should depend only on the probabilities associated with events in the product space $A \times B$, not on the "names" or "identities" of events, and (c) it should satisfy the monotonicity condition that if $p \succsim q$, then at least as much information should be associate with p as q. Property (c) is often called the *data processing inequality*, reflecting the idea that post-processing of data should not provide new information (Cover and Thomas, 2006).

An especially prominent example of an information measure is *mutual information*, defined as the relative entropy between p and the product measure $\bar{p}$ associated with p:[9]

$$MI(p) := R(p \| \bar{p}) = \int_{A \times B} \log\left(\frac{p(a,b)}{p_A(a)p_B(b)}\right) dp(a,b)\,. \tag{1.3}$$

As a function on $\Delta(A \times B)$, mutual information is convex, lower semi-continuous and satisfies properties (a)-(c). Mutual information is only one of many information measures that have been proposed in the information theory literature, but it is the particular information measure that has generally been employed in applications of rational inattention in macroeconomics and finance.

[9] In the definition of mutual information, the standard convention $0 \log(0/0) = 0$ is used.

1.3.2 Components of the representation

In our representation, $\mathcal{S}$ is a given (i.e., primitive) state space. Σ refers to a countably infinite set which we interpret as a latent *signal space*, with typical element (or *signal*) denoted $\sigma \in \Sigma$. As we do not propose to observe *how* the DM processes information, Σ obtains the meaning of a signal space only once there is a joint distribution on the $\mathcal{S}\times\Sigma$ product space. Of course, a joint distribution is not observable either, but our main result shows that a DM satisfying the axioms of Section 1.2 acts *as if* she is choosing a joint distribution on $\mathcal{S}\times\Sigma$, subject to a cost of the information contained in the distribution. The latter is captured by an information cost function.

Definition 1 (Information Cost Function) *Given a prior $\pi \in \Delta(\mathcal{S})$, we say that a convex, lower-semicontinuous function $c : \Delta(\mathcal{S}\times\Sigma) \to [0,\infty]$ is an* information cost function (relative to π) *if c satisfies the following properties:*

(i) Focused on prior π: *$c(p) = \infty$ whenever $p_{\mathcal{S}} \neq \pi$,*

(ii) Grounded at null information: *$c(\bar{p}) = 0$ for all product measures $\bar{p} \in \bar{\Delta}(\mathcal{S}\times\Sigma)$ for which $\bar{p}_{\mathcal{S}} = \pi$,*

(iii) Data processing inequality: *$p \succsim q$ implies $c(p) \geq c(q)$.*

Convexity, lower-semicontinuity, grounded at null information, and the data processing inequality are central properties for information measures in the information theory literature. As such, an information cost function *is* an information measure.

There is one important caveat. Measures of information usually depend only on the probabilities of events, and not the "names" or "identities" of events. An information cost function satisfies this condition only in the signal dimension. Property (iii) implies that cost of attending to information depends only on the probabilities associated to elements of Σ (not their identities). This is natural in our setting as Σ is not a primitive of the decision environment. The interpretation of Σ as a signal space comes only from the probabilities associated to its elements. However, the identities of states are crucial, as the state space is the basic primitive reflecting uncertainty about the real world. In our representation theorem, property (i) of an information cost function therefore constrains the DM to choose joint distributions with $\mathcal{S}$-marginals equal to a unique prior π on the state space, reflecting the primitive status of states.

Finally, we require some notation for utility functions on lotteries. Given a function $u : \mathcal{Z} \to \mathbb{R}$ and a lottery $\mathbb{p} \in \Delta(\mathcal{Z})$, denote by

$$Eu[\mathbb{p}] := \int_{\mathcal{Z}} \mathbb{p}(z)u(z) \,. \tag{1.4}$$

Also, given two functions $u_1, u_2 : \mathcal{Z} \to \mathbb{R}$, we sometimes write $u_1 \approx u_2$ if there exists $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_1 = \alpha\, u_2 + \beta$ (i.e., when u_1 and u_2 are equivalent up to positive affine transformations).

1.3.3 A utility representation for variable attention

We are now able to state our main result, a characterization of preferences satisfying the axioms in Section 1.2.

Theorem 1 *Let $\succsim$ be a binary relation on $\mathcal{M}$. Then, the following two statements are equivalent:*

(i) $\succsim$ satisfies the WO, VA, WSI, MON, MC and UB axioms.

(ii) There exist a prior $\pi \in \Delta(\mathcal{S})$, an unbounded utility index $u : \mathcal{Z} \to \mathbb{R}$, and an information cost function $c : \Delta(\mathcal{S} \times \Sigma) \to [0, \infty]$ (relative to π), such that for all $M, N \in \mathcal{M}$:

$$M \succsim N \quad \Leftrightarrow V(M) \geq V(N) \,, \tag{1.5}$$

where for all $M \in \mathcal{M}$,

$$V(M) = \max_{p \in \Delta(\mathcal{S}\times\Sigma)} \left\{ \int_{\mathcal{S}} \left(\int_{\Sigma} \left[\max_{f \in M} \int_{\mathcal{S}} Eu[f(s')] \, d\, p(s'|\sigma) \right] d\, p(\sigma|s) \right) d\,\pi(s) - c(p) \right\} \tag{1.6}$$

Moreover, if (π_1, u_1, c_1) and (π_2, u_2, c_2) represent the same preference relation $\succsim$ on $\mathcal{M}$, then $\pi_1 = \pi_2$ and there exists $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_1 = \alpha\, u_2 + \beta$ and $c_1 = \alpha\, c_2$.

In view of Theorem 1, we call a preference relation $\succsim$ on $\mathcal{M}$ satisfying axioms WO, VA, WSI, MON, MC and UB, a *preference (relation) with variable attention*; and we call the representation functional in (1.6) a (π, u, c)*-utility function for variable attention* on $\mathcal{M}$.

Proof sketch of Theorem 1. The first step in the sufficiency part of the proof of Theorem 1, is to construct a framework where the objects of choice are simple functions defined on $\mathcal{S}\times\Sigma$. We call these derived objects “superacts”, emphasizing both their artificial nature and act like structure. We then derive a preference relation $\succsim^*$ on superacts through a natural association of superacts with menus. Lemma 1 in the Appendix shows that this derived relation inherits a set of properties once the DM’s preferences $\succsim$ on $\mathcal{M}$ satisfies axioms WO, VA, WSI, MON, MC and UB.

In particular, $\succsim^*$ satisfies all the properties of a variational preference (Maccheroni et al., 2006), except that $\succsim^*$ is ambiguity loving. Adapting Maccheroni et al. (2006), $\succsim^*$ can then be represented by a *convex* niveloid I, that moreover satisfies additional properties derived from the additional structure on $\succsim^*$ (Lemma 2). Lemma 3 identifies the cost function c and demonstrates its properties from the additional structure on the niveloid I. Finally, Lemma 2 and Lemma 3 are used to establish the desired representation. The converse of the result is obtained in a similar fashion as Lemmata 1–3 are characterization results. Details of the proof are given in the Appendix. ■

Theorem 1 shows that the axioms in Section 1.2 characterize a DM who acts *as if* she anticipates choosing joint distributions over $\mathcal{S} \times \Sigma$, subject to the cost of attending to information contained in distributions. The value of a joint distribution $p \in \Delta(\mathcal{S} \times \Sigma)$ comes only from the DM's ability to use the information contained in the distribution to make better *ex-post* choices in the menu.

In particular, since c is focused on π (property (i) of an information cost function), the optimization over joint distributions is constrained so that the DM uses her prior π to form *ex-ante* expectations. This constraint ensures that the DM's prior evaluations are the same for each menu, and varying attention can only improve the choices made *ex-post* (not the *ex-ante* value of a particular plan for choices). If the DM was unable to attend to any information about states before choosing from menus, the best she could do is to choose the best act according to her *a priori* information (captured through π). If, instead, the DM could attend to sufficient information to identify the true state in every case, she could choose the best act in each state in any given menu. However, the DM modeled via (1.6) lies between these extremes (see Observation 1), and optimally varies attention according to the menus under consideration due to the cost of information c. For a DM with variable attention, the value of a menu

M therefore lies somewhere between the expected utility value of the *a priori* optimal act and the *ex-post* optimal act in M.

Observation 1 *Let V be a (π, u, c)-utility function for variable attention. Then, for all $M \in \mathcal{M}$,*

$$\max_{f\in M} \int_{\mathcal{S}} Eu[f(s)]\, d\pi \;\leq\; V(M) \;\leq\; \int_{\mathcal{S}} \max_{f\in M} Eu[f(s)]\, d\pi\,. \tag{1.7}$$

Although we do not characterize specific functional forms for information cost functions, the following proposition establishes that information costs can be measured by the mutual information function, the measure of information proposed by Sims (1998, 2003) for rational inattention modeling.

Proposition 1 *Let $\theta \in [0, \infty]$, $u : \mathcal{Z} \to \mathbb{R}$ be unbounded, $\pi \in \Delta(\mathcal{S})$, and suppose that for all $M \in \mathcal{M}$,*

$$\begin{aligned} V^*(M) = \quad & \max_{p\in\Delta(\mathcal{S}\times\Sigma)} \left\{ \int_{\mathcal{S}} \left(\int_{\Sigma} \left[\max_{f\in M} \int_{\mathcal{S}} Eu[f(s')]\, d\, p(s'|\sigma) \right] d\, p(\sigma|s) \right) d\, p_{\mathcal{S}}(s) \;-\; \theta MI(p) \right\} \\ & s.t. \;\; p_{\mathcal{S}} = \pi\,, \end{aligned} \tag{1.8}$$

then V^* is a utility function with variable attention on $\mathcal{M}$.

1.3.4 A utility representation for constrained attention

In the background of the utility representation for variable attention is the idea that attention is a scarce resource. A special case corresponds to a purely "physical" constraint on attention. Recalling the restaurant choice example, suppose that the DM does not have access to the Internet, newspapers or magazines, but that she could instead call a friend to get a recommendation. If calling a friend has no opportunity cost but she does not have time to call more than one friend, she faces a purely "physical" attention allocation problem.

We obtain a special case of the utility representation for variable attention by strengthening the WSI axiom, which in turn represents a further restriction on the types of preference reversals permitted by VA. Suppose, for example, that the DM has one Italian and one Japanese friend. The restaurant options being considered are Italian and Japanese *à la carte* restaurants, and a *table d'hôte* restaurant. With the physical constraint that she can call only one of her two friends, the optimal allocation of attention for a randomization over the Italian and *table d'hôte* restaurants is not affected by the weight put on the Italian restaurant. No matter how unlikely it is that the Italian restaurant will be drawn, the DM should call her Italian friend for a recommendation. Likewise, in the randomization over the Japanese and *table d'hôte* restaurants, the DM should call her Japanese friend no matter what weight the objective randomization puts on the *à la carte* option. More generally, we argue that under purely "physical" constraints on attention, mixtures with singleton menus should not affect the attention allocation problem. The following axiom therefore posits that they should not lead to preference reversals.

Axiom 7 (Singleton Independence (SI)) *Given* $M, N \in \mathcal{M}$, $f \in \mathcal{F}$ *and* $\alpha \in (0,1)$, *then*

$$\alpha M + (1-\alpha)\{f\} \succsim \alpha N + (1-\alpha)\{f\} \Leftrightarrow M \succsim N \ .$$

As the intuition behind the restaurant choice example suggests, strengthening WSI to SI leads to an *as if* model of an attention allocation problem in which the DM does not face costs of attention, but instead optimizes over joint distributions in a constraint set. The following definition formalizes the properties which the resulting constraint set satisfies.

Definition 2 (Information Constraint Set) *Given a prior $\pi \in \Delta(\mathcal{S})$, we say that a compact convex set $\Pi \subset \Delta(\mathcal{S} \times \Sigma)$ is an* information constraint set (relative to π) *if Π satisfies the following conditions:*

(i) Focused on prior π: $p_{\mathcal{S}} \neq \pi$ implies $p \notin \Pi$,

(ii) Inclusive of null information: $\{p \in \bar{\Delta}(\mathcal{S} \times \Sigma) | p_{\mathcal{S}} = \pi\} \subset \Pi$,

(iii) Data processing monotonicity: $p \in \Pi$ and $p \succsim q$ implies $q \in \Pi$.

Properties (i)-(iii) of an information constraint set are the natural counterparts of properties (i)-(iii) of an information cost function, and have the same interpretation in the current setting. The focus on π restricts the DM to use the same prior to form all *ex-ante* expectations. The inclusion of null information implies that an information constraint set is non-empty, because it contains joint distributions under which $\mathcal{S}$ and Σ are independent. Data processing monotonicity implies that an information constraint set that contains p also contains any distributions q that contain less information than p. The following representation result characterizes when attention of a preference relation can be captured through an information constraint set.

Theorem 2 *Let $\succsim$ be a binary relation on $\mathcal{M}$. Then, the following two statements are equivalent:*

(i) $\succsim$ is a preference relation with variable attention that satisfies the SI axiom.

(ii) There exists a prior $\pi \in \Delta(\mathcal{S})$, an unbounded utility index $u : \mathcal{Z} \to \mathbb{R}$, and an information constraint set $\Pi \subset \Delta(\mathcal{S}\times\Sigma)$ (relative to π), such that for all $M, N \in \mathcal{M}$,

$$M \succsim N \;\Leftrightarrow\; \tilde{V}(M) \geq \tilde{V}(N)\,, \tag{1.9}$$

where, for all $M \in \mathcal{M}$,

$$\tilde{V}(M) = \max_{p\in\Pi} \int_{\mathcal{S}} \left(\int_{\Sigma} \left[\max_{f\in M} \int_{\mathcal{S}} Eu[f(s')]\, d\, p(s'|\sigma) \right] d\, p(\sigma|s) \right) d\,\pi(s)\,. \tag{1.10}$$

Moreover, if (π_1, u_1, Π_1) and (π_2, u_2, Π_2) represent the same preference relation $\succsim$ on $\mathcal{M}$, then $\pi_1 = \pi_2$, $\Pi_1 = \Pi_2$, and $u_1 \approx u_2$.

It is clear that WSI is redundant in the presence of SI. Hence, in view of Theorem (2), we call a binary relation $\succsim$ on $\mathcal{M}$ satisfying axioms WO, VA, SI, MON, MC and UB, a *preference (relation) with constrained attention*; and we call the representation functional in (1.10) a (π, u, Π)-*utility representation with constrained attention* on $\mathcal{M}$. Proposition 2 confirms that the information constraint set can again be defined in terms of the mutual information function. The resulting model is the finite Shannon capacity functional form first proposed by Sims (1998) for rational inattention modeling.

Proposition 2 *Let $\kappa \in [0, \infty]$, $u : \mathcal{Z} \to \mathbb{R}$ unbounded, $\pi \in \Delta(\mathcal{S})$, and suppose that for all $M \in \mathcal{M}$*

$$\begin{aligned} \tilde{V}^*(M) = \quad & \max_{p\in\Delta(\mathcal{S}\times\Sigma)} \int_{\mathcal{S}} \left(\int_{\Sigma} \left[\max_{f\in M} \int_{\mathcal{S}} Eu[f(s')]\, d\, p(s'|\sigma) \right] d\, p(\sigma|s) \right) d\, p_{\mathcal{S}}(s) \\ & s.t. \;\; p_{\mathcal{S}} = \pi \;\text{ and }\; MI(p) \leq \kappa, \end{aligned} \tag{1.11}$$

then $\tilde{V}^*$ is a utility function with constrained attention on $\mathcal{M}$.

1.3.5 Fixed attention and subjective learning

A recent literature studies models of subjective learning (Dillinberger and Sadowski, 2011; Takeoka, 2005; Lleras, 2010) in similar decision environments to the one outlined in Section 1.2. Subjective learning models study a DM who anticipates receiving subjective information – i.e, information not observable to a decision theorist – before making *ex-post* choices, and try to capture the behavioral implications of expected information gain. There are noticeable intersections with our decision theoretic model of scarce attention, which we now formalize.

Unlike an agent with scarce attention, the DM in the subjective learning literature is a *passive* recipient of information, and is neither able to exert more (or less) effort to change her knowledge about states, nor able to actively tailor the information she attends to for the particular menu under consideration. In the context of the restaurant choice example, one could think of a DM who anticipates meeting a particular colleague on the way to a restaurant, anticipates getting information from the colleague that could help in choosing a dish, but does not anticipate being able to actively influence the information that the colleague will reveal to her. In some sense, the DM anticipates learning information that a decision theorist can not observe, but the DM does not in fact face any attention allocation problem.

To clarify the relation between our model of scarce attention and models of subjective learning formally, we provide a further special case of the representations in Theorems 1 and 2.[10] We obtain a model of passive learning simply by

[10]Axiomatically, our characterization of subjective learning adapts the results in Dillinberger and Sadowski (2011) Takeoka (2005, also closely related to) to our modified setting, but our representation of expected information in terms of a joint distribution on a product space of states with latent signals is distinct (and follows a different method of proof).

strengthening the VA axiom.

Our normative justification for VA was based on the idea that the DM would ideally prefer the resolution of objective randomization *before* she has actively allocated attention among competing information sources, as late resolution could complicate the attention allocation problem. However, if the DM anticipates being a passive recipient of information (i.e., does not believe that she will be able to actively vary the information that she can attend to), then there is no reason to posit that she might favor early resolution of objective uncertainty (at least not, from the informational viewpoint that justified VA). As such, the following *fixed attention* axiom states that a DM indifferent between two menus will also be indifferent to any randomization over the menus.

Axiom 8 (Fixed Attention (FA)) *Given $M, N \in \mathcal{M}$, then $M \sim N$ implies $M \sim \alpha M + (1-\alpha)N$ for all $\alpha \in (0,1)$.*

A representation for subjective learning is characterized in the context of our other axioms by strengthening VA to FA.

Theorem 3 *Let $\succsim$ be a binary relation on $\mathcal{M}$. Then, the following two statements are equivalent:*

(i) $\succsim$ is a preference relation with variable attention that satisfies FA.

(ii) There exist a prior $\pi \in \Delta(\mathcal{S})$, an unbounded utility index $u : \mathcal{Z} \to \mathbb{R}$, and a joint distribution $p \in \Delta(\mathcal{S} \times \Sigma)$ with $p_{\mathcal{S}} = \pi$, such that for all $M, N \in \mathcal{M}$:

$$M \succsim N \quad \Leftrightarrow \check{V}(M) \geq \check{V}(N)\,, \tag{1.12}$$

where for all $M \in \mathcal{M}$,

$$\check{V}(M) = \int_{\mathcal{S}} \left(\int_{\Sigma} \left[\max_{f \in M} \int_{\mathcal{S}} Eu[f(s')] \, d\, p(s'|\sigma) \right] d\, p(\sigma|s) \right) d\, \pi(s) \,. \tag{1.13}$$

Moreover, if (π_1, u_1, p_1) and (π_2, u_2, p_2) represent the same preference relation $\succsim$ on $\mathcal{M}$, then $\pi_1 = \pi_2$, $u_1 \approx u_2$, and $p_1 \simeq p_2$.

In view of Theorem (3), we call a binary relation $\succsim$ on $\mathcal{M}$ satisfying axioms WO, FA, WSI, MON, MC and UB, a *preference (relation) with fixed attention*; and we call the representation functional in (1.13) a (π, u, p)-*utility function with fixed attention* on $\mathcal{M}$.

The fixed attention model in general allows for a DM to be inattentive to some available information, but the model differs from the variable and constrained attention models by placing further restrictions on the type of (in)attention that the DM can exhibit. The DM now acts *as if* she anticipates attending to certain information about states before she will choose from menus, but receives this information passively (i.e., is not able to optimize information gain with respect to the particular menu under consideration). It seems clear to us that VA is normatively more compelling than FA, but as a descriptive model, fixed attention provides a behavioral model of subjective learning that identifies the fixed constraint on anticipated information gain from choice data.

The uniqueness properties of the joint distribution $p \in \Delta(\mathcal{S} \times \Sigma)$ is crucial for the behavioral interpretation of the representation of preferences with fixed attention. The joint distribution used by the DM to choose from and evaluate menus in a utility function with fixed attention is distinguished by the prior π over the state space, and the informational content which signals convey about realized states *ex-post*. As Theorem 3 demonstrates, both are unique.

As such, Theorem 3 characterizes when the information that a DM expects to receive about realized states (without any ability to actively tailor the information attended to) can be uniquely identified from choice behavior in our setting; namely, exactly when the DM has preferences with fixed attention.

1.4 Comparative Attention

In this Section, we formalize a notion of comparative attention. We first require some notation. Let $\mathcal{P}$ denote the set of all preference relations on $\mathcal{M}$ with variable attention.[11] For any $\succsim \in \mathcal{P}$, denote the set of preferences with variable attention that coincide with $\succsim$ on singleton menus by $\mathcal{P}(\succsim) = \{ \succsim' \in \mathcal{P} : \{f\} \succsim' \{g\}$ iff $\{f\} \succsim \{g\}$ for all $f, g \in \mathcal{F}\}$.

Observation 2 *Let $\succsim_1, \succsim_2 \in \mathcal{P}$ be represented by (u_1, π_1, c_1) and (u_2, π_2, c_2), respectively. Then the following statements are equivalent.*

(i) $\succsim_1, \succsim_2 \in \mathcal{P}(\succsim)$ *for some* $\succsim \in \mathcal{P}$.

(ii) $u_1 \approx u_2$ *and* $\pi_1 = \pi_2$.

For two DMs with variable attention, different preferences over menus can arise due to different tastes for prizes (u), different priors (π), or different costs of information (c). Observation 2 demonstrates that if their preferences on singleton menus coincide, their tastes and priors also coincide. We then seek a behavioral criterion to relate their information costs, and thereby to establish

[11]Formally, $\mathcal{P} := \{\succsim \subset \mathcal{M} \times \mathcal{M} : \succsim$ satisfies WO, VA, WSI, MON, MC and U$\}$.

observational implications of the (in)attention exhibited by preferences.[12]

To motivate, consider two DMs, Ann and Bob, who reveal the same preferences over *table d'hôte* restaurants. Moreover, suppose that whenever Ann prefers an à la carte restaurant over a specific *table d'hôte* restaurant, Bob reveals the same preference. We would like to infer from their preferences over *table d'hôte* restaurants that Ann and Bob have similar tastes for dishes, and from Bob's greater willingness to dine at *à la carte* restaurants that he expects to be able to make more informed choices from any menu (i.e., that Bob is more attentive to information than Ann). Our comparative notion of (in)attention formalizes this idea.

Definition 3 (More Attentive) *If $\succsim_1, \succsim_2 \in \mathcal{P}$, we say that $\succsim_2$ is* more attentive *than $\succsim_1$ if, for all $M \in \mathcal{M}$ and $f \in \mathcal{F}$,*

$$M \succsim_1 \{f\} \;\Rightarrow\; M \succsim_2 \{f\}\,. \tag{1.14}$$

Proposition 3 *Let $\succsim_1, \succsim_2 \in \mathcal{P}$ be represented by (π_1, u_1, c_1) and (π_2, u_2, c_2), respectively. Then the following two statements are equivalent:*

(i) $\succsim_2$ *is more attentive than* $\succsim_1$.

(ii) $\pi_1 = \pi_2$, $u_1 \approx u_2$, *and* $c_1 \geq c_2$ *(when $u_1 = u_2$).*

Proposition 3 confirms that the behavioral criterion for attentiveness in Definition 3 is characterized in terms of intuitive comparative statics on the costs of information. In particular, if $\succsim_1$ and $\succsim_2$ are represented by (π_1, u_1, θ_1) and

[12]Our behavioral criterion for (in)attention are similar in spirit to the comparative notion of ambiguity aversion proposed in Ghirardato and Marinacci (2002), but adapted to a different setting.

(π_2, u_2, θ_2), respectively, as in the representation of Proposition 1, then $\succsim_2$ is more attentive than $\succsim_1$ if and only if $\pi_1 = \pi_2$, $u_1 \approx u_2$, and $\theta_1 \geq \theta_2$ (when $u_1 = u_2$). The same behavioral criterion for (in)attention allows for equally intuitive comparative statics in the constrained and fixed attention models.

Corollary 1 *Let $\succsim_1, \succsim_2 \in \mathcal{P}$ be preference relations with constrained [fixed] attention associated with information consraint sets [fixed information points], Π_1 $[p_1]$ and Π_2 $[p_2]$, respectively. Then the following two statements are equivalent:*

(i) $\succsim_2$ *is more attentive than* $\succsim_1$*.*

(ii) $\succsim_1, \succsim_2 \in \mathcal{P}(\succsim)$ *for some* $\succsim \in \mathcal{P}$*, and* $\Pi_1 \subset \Pi_2$ $[p_2 \succsim p_1]$.

In particular, if $\succsim_1$ and $\succsim_2$ on $\mathcal{M}$ are represented by (π_1, u_1, κ_1) and (π_2, u_2, κ_2) as in the representation of Proposition 2, respectively, then $\succsim_2$ is more attentive than $\succsim_1$ if and only if $\pi_1 = \pi_2$, $u_1 \approx u_2$, and $\kappa_1 \leq \kappa_2$.

Based on the comparative notion of (in)attention in Definition 3, we can also define absolute notions of (in)attention to describe, respectively, the behavior of most and least attentive DMs. To this end, we say that a preference relation $\succsim \in \mathcal{P}$ is *totally attentive* if it is more attentive than every $\succsim' \in \mathcal{P}(\succsim)$; and we say that $\succsim$ is *totally inattentive* if every $\succsim' \in \mathcal{P}(\succsim)$ is more attentive than $\succsim$. Formally, these notions of absolute (in)attention lead us to the following definition.

Definition 4 (Totally (In)Attentive) *A preference relation* $\succsim \in \mathcal{P}$ *is:*

(i) totally attentive *if, for all* $\succsim' \in \mathcal{P}(\succsim)$*,*

$$M \succsim' \{f\} \Rightarrow M \succsim \{f\} \text{ for all } M \in \mathcal{M} \text{ and } f \in \mathcal{F}\,. \tag{1.15}$$

(ii) totally inattentive *if, for all* $\succsim' \in \mathcal{P}(\succsim)$,

$$M \succsim \{f\} \Rightarrow M \succsim' \{f\} \text{ for all } M \in \mathcal{M} \text{ and } f \in \mathcal{F}\,. \tag{1.16}$$

Intuitively, in each state of the world, a totally attentive DM is able to observe the true state without costs (i.e., has unlimited attention resources). A totally inattentive DM meanwhile, is not able to attend to any information after the realization of states, and can therefore make *ex-post* choices based only on her *a priori* information. These two models represent extremes in terms of the ability to attend to information about states, but both clearly exhibit fixed attention as, for quite different reasons, neither faces any trade-offs in attention allocation. In the following Propositions we characterize both attention extremes in terms of two additional behavioral axioms.

Axiom 9 (Strong Monotonicity (SMON)) *Given* $M, N \in \mathcal{M}$ *such that for all* $g \in N$ *and* $s \in \mathcal{S}$ *there exists* $f \in M$ *satisfying* $\{f(s)\} \succsim \{g(s)\}$, *then* $M \succsim N$.

Axiom 10 (Null Attention (NA)) *Given* $M, N \in \mathcal{M}$, $M \succsim N$ *implies* $M \sim M \cup N$.

Both SMON and NA can be viewed as monotonicity conditions on preferences. SMON is stronger than MON, as for all $g \in N$ if there exists an $f \in M$ that is preferred to g in *all* states, then the premise of SMON is clearly satisfied. It is also easily observed that NA implies MON when combined with the usual monotonicity condition on singleton menus (see p. 9).

To state characterization results for preferences that satisfy SMON or NA, we require some additional notation. For any $\pi \in \Delta(\mathcal{S})$, we denote by $\Delta_\pi(\mathcal{S} \times$

$\Sigma) := \{p \in \Delta(\mathcal{S} \times \Sigma) : p_{\mathcal{S}} = \pi\}$ the set of all joint distributions on $\mathcal{S} \times \Sigma$ with $\mathcal{S}$-marginal π, and we denote by $\bar{\Delta}_\pi := \left\{p \in \bar{\Delta}(\mathcal{S} \times \Sigma) : p_{\mathcal{S}} = \pi\right\}$ the set of all product distributions on $\mathcal{S}\times\Sigma$ with $\mathcal{S}$-marginal π. Finally, for any $p \in \Delta(\mathcal{S}\times\Sigma)$, we denote by $\Lambda(p) = \{q \in \Delta(\mathcal{S} \times \Sigma) : p \succsim q\}$ the set of all joint distributions on $\mathcal{S} \times \Sigma$ for which p is sufficient.

Proposition 4 *Let $\succsim \in \mathcal{P}$ be represented by the triple (π, u, c). Then the following statements are equivalent:*

(i) $\succsim$ *satisfies SMON.*

(ii) $V(M) = \int_S \left(\max_{f\in M} Eu[f(s)]\right) d\pi(s)$ *for all* $M \in \mathcal{M}$.

(iii) $c(p) = \begin{cases} 0 & \textit{if} \quad p \in \Delta_\pi(S \times \Sigma) \\ \infty & \quad \textit{otherwise} \end{cases}$

(iv) $\succsim$ *is represented by the triple* (π, u, Π) *with* $\Pi = \Delta_\pi(\mathcal{S} \times \Sigma)$.

(v) $\succsim$ *is represented by the triple* (π, u, p) *with* $\Lambda(p) = \Delta_\pi(\mathcal{S} \times \Sigma)$.

(vi) $\succsim$ *is totally attentive.*

Proposition 4 characterizes total attention in terms of SMON; our final result gives the characterization for preferences that are totally inattentive in terms of NA.

Proposition 5 *Let $\succsim \in \mathcal{P}$ be represented by the triple (π, u, c). Then the following statements are equivalent:*

(i) $\succsim$ *satisfies NA.*

(ii) $V(M) = \max_{f\in M} \int_S Eu[f(s)] d\pi(s)$ *for all* $M \in \mathcal{M}$.

(iii) $c(p) = \begin{cases} 0 & \text{if } p \in \bar{\Delta}_\pi(\mathcal{S} \times \Sigma) \\ \infty & \text{otherwise} \end{cases}$

(iv) $\succsim$ *is represented by the triple* (π, u, Π) *with* $\Pi = \bar{\Delta}_\pi(\mathcal{S} \times \Sigma)$.

(v) $\succsim$ *is represented by the triple* (π, u, p) *with* $\Lambda(p) = \bar{\Delta}_\pi(\mathcal{S} \times \Sigma)$.

(vi) $\succsim$ *is totally inattentive.*

1.5 Discussion of Related Literature

Part of the motivation for our analysis comes from the many applications of Sims' rational inattention model (see, e.g., Wiederholt, 2010; Veldkamp, 2011, for a review and references). Mutual information is the standard information measure used in this literature. Although we do not characterize specific functional forms of information cost functions and constraint sets, Propositions 1 and 2 demonstrate that the most commonly used objective functions in the rational inattention literature are consistent with the utility functionals we axiomatize. Our results therefore complement the literature applying rational inattention models in macroeconomics and finance by providing behavioral foundations that help to clarify their observational content.

In addition, several strands of the decision theory literature inform our approach. The axioms that we use to characterize decision making with scarce attention are closely related to axioms commonly found in the literature on ambiguity aversion on Anscombe-Aumann choice domains, although the interpretation in a menu choice setting is necessarily quite different. For example, the proof for Theorem 1 clearly relates the variable attention model to the variational preferences introduced in Maccheroni et al. (2006), a model that generalizes maxmin expected utility (Gilboa and Schmeidler, 1989) in a similar way as variable attention generalizes constrained attention. Maccheroni et al. (2006) offer a game-against-nature interpretation for the variational utility representation, in which the DM acts *as if* she faces a malevolent nature that, at a cost, chooses distributions over the state space against her. Although our models of scarce attention have related axiomatic foundations, the interpretation offered is quite different. In our models, it is the DM who chooses distributions and faces subjective costs specifically from attending to information. The DM is otherwise a subjective expected utility maximizer – the epitome of a Bayesian DM – and deviations from the "standard" rational expectations model come only from limitations on resources for information processing.

It is important for our information-theoretic interpretations that we identify more behavioral primitives and properties of the cost function than Maccheroni et al. (2006), and we are able to do so by exploiting our richer menu choice domain. Menu choice problems were first proposed by Kreps (1979) to study preference for flexibility. An influential paper in the subsequent literature is Dekel et al. (2001), who use menus over objective lotteries on a set of prizes to characterize a subjective state space representation. In our domain on menus of acts, the state space is a primitive but the signal space – which is derived from preferences – could be viewed as subjective. However, the interpretation is quite different. The meaning of the state space in the variable attention model comes from the joint distribution *chosen* by the DM, and depends only on the probabilities of different signals not their identity (by property (iii) of an information cost function). As in the rational inattention approach more generally, it is not necessary to try and interpret the signal space "independently" of some beliefs over it, as Dekel et al. (2001) do in their subjective state space model. The only probability measure that needs to be interpreted independently is the marginal on the primitive state space $\mathcal{S}$, and the uniqueness of π in our representation results assures that we are able to do so.

There are also related menu choice models in which DMs face subjective costs from, *inter alia*, self-control problems (e.g., Gul and Pesendorfer, 2001) or thinking aversion (Ortoleva, 2012). A driving assumption in many of these models is that DMs may sometimes strictly prefer a subset of a menu (unlike in Kreps, 1979). Ortoleva (2012), for example, studies a DM who incurs costs from thinking about the content of a menu, and may therefore prefer a smaller menu to a larger one that contains it. Masatlioglu et al. (2012) also consider a DM who reveals potentially limited attention about items in a menu. Thinking aversion and the revealed inattention in Masatlioglu et al. (2012) differ, however, from the type of inattention we study because a DM with variable attention fully understands the content of each menu, and rather faces costs from processing information about an external state space that is independent of menus. A strict preference for subsets therefore does not make sense in our setting – or, at least, does not follow from the type of inattention we are interested in – and it is straightforward to verify that our monotonicity axiom contradicts such preference patterns.

The closest work to ours in the menu choice literature is Ergin and Sarver (2010), who extend on Dekel et al. (2001) to analyze a decision model with costly contemplation.[13] In the costly contemplation model, the DM has uncertainty

[13]After completing our manuscript, we became aware of independent, similar contributions by Ellis (2012), and (via private communications) by Denti (2012) and de Oliveira (2012). Ellis (2012) also presents a decision model motivated by rational inattention, but in terms of decision environment, axioms and representation his approach is very different to ours. The works of Denti (2012) and de Oliveira (2012) seem close to ours and both contain a representation for what we call preferences with variable attention in this paper. However, their methods of proof differ substantially from ours, and we have seen only an early draft of their papers, that do not

about future tastes (e.g., parameters of risk-aversion) and, at a subjective cost, can exert contemplation to better understand her tastes before choosing from the menu. The domain of preferences in their analysis is menus of objective lotteries, and therefore differs from ours. But the axioms that characterize costly contemplation behaviorally are very similar to ours. In particular, our variable attention axiom is the natural counterpart of their contemplation aversion axiom in a richer choice setting. However, the translation of their axioms into a menu-of-acts domain leads to new insights. A DM with variable attention understands her tastes, and uncertainty about external states remains only due to limitations on her information processing. Our representation is therefore closer to applied work in information economics, as testified by the large preexisting literature in which rational inattention has been successfully used to explain diverse phenomena from price stickiness (e.g., Mackowiak and Wiederholt, 2009) and business cycle fluctuations (e.g., Mackowiak and Wiederholt, 2010), to the home-bias in international portfolio decisions (e.g., Van Nieuwerburgh and Veldkamp, 2009).

Finally, the decision models we study could be placed in the context of the large literature on bounded rationality more generally, although that label seems misleading to us. To be sure, rational inattention is a deviation from the rational expectations framework, where an agent is assumed to internalize all publicly available information without effort. But the manner in which information processing is treated under rational expectations seem like an unreasonably high bar to place on rationality. As cognitive psychologists have long recognized, attention can be viewed as a scarce resource, and optimal allocation of attention could then be viewed a classic resource allocation problem. A central

represent alternative forms of (in)attention, as in our fixed attention.

contribution of Sims (1998, 2003) was to draw this resource allocation problem to the attention of the economic profession, and propose a tractable framework to integrate attention allocation into larger economic models. A growing literature attests to the value of his proposed approach, but we believe that the study of inattention also falls in the purview of decision theory as the allocation of attention – unlike many other valuable resources – often occurs in private, behind closed doors, rather than in markets or other public forums. Firmly in the spirit of Sims' approach, we nevertheless regard attention as the allocation of a scarce resource by an optimizing agent, and therefore hold that as a model of behavior our findings affirm that rational inattention has both descriptive and normative appeal.

1.6 Conclusion

The objective functions used in economic models are chosen to capture pertinent behavioral attributes of agents in a tractable way. Rational inattention is a case in point. The assumption that an agent chooses a joint distribution over states and signals omits many details of attention and information processing, but has proven tractable enough to successfully embed attention allocation problems in larger economic models. In order to remain empirically founded it is also important, however, to understand what the falsifiable implications of a decision rule are. The objects of choice in the rational inattention approach (i.e., joint distributions) make this task appear challenging.

In this paper, we allay concerns over the observational foundations of rational inattention models by identifying a choice domain and a set of falsifiable axioms that characterize rationally inattentive behavior. A DM who obeys our axioms acts *as if* she anticipates choosing joint distributions over states and signals subject to costs of information, as in much of the applied literature on rational inattention. The attention allocation problem proposed by Sims (1998, 2003) therefore attains the same standing in our framework as the prior in the classic subjective expected utility model of Savage (1954), itself not observable but inferable from observable choice data.

As a large literature attests, rational inattention appears to have much to offer for economic research in the age of information. By clarifying some of the decision theoretic underpinnings of rational inattention, our results therefore add to a deeper understanding of an important approach to model limitations on people's ability to attend to the abundance of available information.

1.7 Appendix

1.7.1 Preliminaries

We first introduce some notation and definitions extending on Maccheroni et al. (2006).

For any finite set $\mathcal{S}$ and countably infinite set Σ, let $B_0(\mathcal{S} \times \Sigma)$ be the set of all simple functions $\varphi : \mathcal{S} \times \Sigma \to \mathbb{R}$, and $B(\mathcal{S} \times \Sigma)$ its supnorm closure consisting of all bounded functions $\psi : \mathcal{S} \times \Sigma \to \mathbb{R}$. $B_0(\mathcal{S} \times \Sigma, T)$ denotes the set of all functions in $B_0(\mathcal{S} \times \Sigma)$ taking values in the interval $T \subset \mathbb{R}$. Endowed with the supnorm, $B_0(\mathcal{S} \times \Sigma)$ is a normed vector space and its norm dual $ba(\mathcal{S} \times \Sigma)$ is endowed with the total variation norm. As a consequence of duality, the space of all probability measures $\Delta(S \times \Sigma)$ can be endowed with the weak* topology where a net $\{p_d\}_{d\in D}$ converges to p if $p_d(E) \to p(E)$ for all $E \in 2^{\mathcal{S}\times\Sigma}$ (see, e.g., the opening remarks in the Appendix of Maccheroni et al., 2006).

For $\varphi, \psi \in B(\mathcal{S}\times\Sigma)$, we write $\varphi \geq \psi$ if $\varphi(s,\sigma) \geq \psi(s,\sigma)$ for all $(s,\sigma) \in \mathcal{S}\times\Sigma$, and we write $\varphi \geq_\Sigma \psi$ if for all $\sigma \in \Sigma$ there exists $\sigma' \in \Sigma$ such that $\varphi(s,\sigma') \geq \psi(s,\sigma)$ for all $s \in \mathcal{S}$. Let Φ be any non-empty, convex collection of elements of $B(\mathcal{S}\times\Sigma)$, Φ_c the subset of constant functions in Φ, and Φ_S the subset of functions in Φ that are measurable with respect to the algebra $2^S \times \Sigma$. Given a functional $I : \Phi \to \mathbb{R}$, we say that I is:

(i) *normalized* if $I(k\mathbb{1}_{\mathcal{S}\times\Sigma}) = k$ for all $k\mathbb{1}_{\mathcal{S}\times\Sigma} \in \Phi_c$;

(ii) *monotonic* if $\varphi, \psi \in \Phi$ and $\varphi \geq \psi$ implies $I(\varphi) \geq I(\psi)$;

(iii) Σ-*monotonic* if $\varphi, \psi \in \Phi$ and $\varphi \geq_\Sigma \psi$ implies $I(\varphi) \geq I(\psi)$;

(iv) *singleton-additive* if $I(\alpha\varphi + (1-\alpha)\psi) = I(\alpha\varphi) + (1-\alpha)I(\psi)$ for all $\varphi \in \Phi$, $\psi \in \Phi_S$ and $\alpha \in [0,1]$ such that $\alpha\varphi \in \Phi$;

(v) *convex* if $I(\alpha\varphi+(1-\alpha)\psi) \leq \alpha I(\varphi)+(1-\alpha)I(\psi)$ for all $\varphi, \psi \in \Phi$ and all $\alpha \in [0,1]$ such that $\alpha\varphi + (1-\alpha)\psi \in \Phi$;

(vi) a *niveloid* if $I(\varphi) - I(\psi) \leq \sup(\varphi - \psi)$ for all $\varphi, \psi \in \Phi$.

Remark 1 *Let $\bar{I} : -\Phi \to \mathbb{R}$ be defined by $\bar{I}(\varphi) := -I(-\varphi)$ for all $\varphi \in -\Phi$. Then $\bar{I}$ is normalized iff I is normalized; $\bar{I}$ is monotonic iff I is monotonic; $\bar{I}$ is Σ-monotonic iff I is Σ-monotonic; $\bar{I}$ is singleton-additive iff I is singleton-additive; $\bar{I}$ is a niveloid iff I is a niveloid; $\bar{I}$ is* concave *iff I is* convex.

1.7.2 "Superacts"

Let $\mathcal{K}_0$ denote the set of all simple functions $K : \mathcal{S} \times \Sigma \to \Delta(\mathcal{Z})$. We refer to elements of $\mathcal{K}_0$ as *superacts*. Denote by $\mathcal{K}_c$ the set of all $K_{\mathbb{p}} \in \mathcal{K}_0$ – called *constant superacts* – for which there exists a $\mathbb{p} \in \Delta(\mathcal{Z})$ such that $K_{\mathbb{p}}(s, \sigma) = \mathbb{p}$ for all $(s, \sigma) \in \mathcal{S} \times \Sigma$. Denote by $\mathcal{K}_S$ the set of all $K_f \in \mathcal{K}_0$ – called *singleton superacts* – for which there exists an act $f \in \mathcal{F}$ such that $K_f(., \sigma) = f$ for all $\sigma \in \Sigma$. With some abuse of notation, we denote by $K(s, \sigma) \in \mathcal{K}_0$ the super act that takes value $K(s, \sigma) \in \Delta(\mathcal{Z})$ for all $(s', \sigma') \in \mathcal{S} \times \Sigma$. For all $K, K' \in \mathcal{K}_0$ and $\alpha \in [0, 1]$, $\alpha K + (1 - \alpha)K'$ is the superact $\bar{K} \in \mathcal{K}_0$ defined by $\bar{K}(s, \sigma) := \alpha K(s, \sigma) + (1 - \alpha)K'(s, \sigma)$ for all $(s, \sigma) \in \mathcal{S} \times \Sigma$. Note that with respect to this mixture operation, the set $\mathcal{K}_0$ is a mixture space.

In the sequel, we will refer to the following properties of a binary relation $\succsim^\star$ on $\mathcal{K}_0$. To this end, associate $\succsim^\star$ in the usual sense, with its symmetric and asymmetric parts denoted $\sim^\star$ and $\succ^\star$, respectively.

- *Weak Order (WO*):* (i) If $K, K' \in \mathcal{K}_0$, then either $K \succsim^\star K'$ or $K' \succsim^\star K$. (ii) If $K, K', K'' \in \mathcal{K}_0$, $K \succsim^\star K'$ and $K' \succsim^\star K''$, then $K \succsim^\star K''$.
- *Ambiguity Affection (AA*):* If $K, K' \in \mathcal{K}_0$, $K \sim^\star K'$ and $\alpha \in (0, 1)$, then $K \succsim^\star \alpha K + (1 - \alpha)K'$.
- *Weak Constant-Singleton Independence (WCSI*):* If $K, K' \in \mathcal{K}_0$, $K_{\mathbb{p}}, K_{\mathbb{q}} \in \mathcal{K}_c$ and $\alpha \in (0, 1)$, then

$$\alpha K + (1 - \alpha)K_{\mathbb{p}} \succsim^\star \alpha K' + (1 - \alpha)K_{\mathbb{p}} \;\Leftrightarrow\; \alpha K + (1 - \alpha)K_{\mathbb{q}} \succsim^\star \alpha K' + (1 - \alpha)K_{\mathbb{q}} \,.$$

- *Weak Singleton Independence (WSI*):* If $K, K' \in \mathcal{K}_0$, $K_f, K_g \in \mathcal{K}_S$ and $\alpha \in (0, 1)$, then

$$\alpha K + (1 - \alpha)K_f \succsim^\star \alpha K' + (1 - \alpha)K_f \;\Leftrightarrow\; \alpha K + (1 - \alpha)K_g \succsim^\star \alpha K' + (1 - \alpha)K_g \,.$$

- *Singleton Independence (SI*):* If $K, K' \in \mathcal{K}_0$, $K_f \in \mathcal{K}_S$ and $\alpha \in (0,1)$, then

$$K \succsim^\star K' \;\Leftrightarrow\; \alpha K + (1-\alpha)K_f \succsim^\star \alpha K' + (1-\alpha)K_f \ .$$

- *Monotonicity (MON*):* If $K, K' \in \mathcal{K}_0$ such that $K(s,\sigma) \succsim^\star K'(s,\sigma)$ for all $(s,\sigma) \in \mathcal{S} \times \Sigma$, then $K \succsim^\star K'$.

- Σ*-Monotonicity (*Σ*MON*):* If $K, K' \in \mathcal{K}_0$ such that for all $\sigma \in \Sigma$, there exists some $\sigma' \in \Sigma$ with $K(s,\sigma') \succsim^\star K'(s,\sigma)$ for all $s \in \mathcal{S}$, then $K \succsim^\star K'$.

- *Mixture-Continuity (MC*):* For all $K_f, K_g \in \mathcal{K}_S$ and $K \in \mathcal{K}_0$, the sets $\{\alpha \in [0,1] \mid \alpha K_f + (1-\alpha)K_g \succsim^\star K\}$ and $\{\alpha \in [0,1] \mid K \succsim^\star \alpha K_f + (1-\alpha)K_g\}$ are closed.

- *Unboundedness (UB*):* There exist $K_{\mathbb{p}}, K_{\mathbb{q}} \in \mathcal{K}_c$ with $K_{\mathbb{p}} \succ^\star K_{\mathbb{q}}$ such that for all $\alpha \in (0,1)$ there exists $K_{\mathbb{r}} \in \mathcal{K}_c$ satisfying either $\alpha K_{\mathbb{r}} + (1-\alpha)K_{\mathbb{q}} \succ^\star K_{\mathbb{p}}$ or $K_{\mathbb{q}} \succ^\star \alpha K_{\mathbb{r}} + (1-\alpha)K_{\mathbb{p}}$.

Properties of a derived relation on superacts

The focus of this Section is a Lemma that characterizes properties of a derived binary relation on $\mathcal{K}_0$. Before stating the result, we require some additional notation and definitions.

Let $\{\sigma_1, \sigma_2, ...\}$ be an enumeration of the set Σ, and for any $M = \{f_1, ..., f_{T_M}\} \in \mathcal{M}$, define $K_M \in \mathcal{K}_0$ as follows:

$$K_M(.,\sigma_j) = \begin{cases} f_j & \text{if } j \leq T_M \\ f_1 & \text{if } j > T_M \end{cases} .$$

In all of the following, the specific enumerations chosen for M and Σ do not matter.

Define the operator $m : \mathcal{K}_0 \to \mathcal{M}$ as follows: For all $K \in \mathcal{K}_0$ let

$$m(K) = \{f \in \mathcal{F} \mid f = K(., \sigma) \text{ for some } \sigma \in \Sigma\} \ . \tag{1.17}$$

Remark 2 *Let $K, K' \in \mathcal{K}_0$, $M \in \mathcal{M}$, $f \in \mathcal{F}$, $K_f \in \mathcal{K}_S$, $\alpha \in [0, 1]$. Then,*

(i) $m(K) \in \mathcal{M}$, as for each $\sigma \in \Sigma$ $K(., \sigma) \in \mathcal{F}$ and the range of K is finite.

(ii) $m(K_M) = M$ and $m(K_f) = \{f\}$.

(iii) $m(\alpha K + (1 - \alpha)K') \subset \alpha m(K) + (1 - \alpha)m(K')$.

(iv) $m(\alpha K + (1 - \alpha)K_f) = \alpha m(K) + (1 - \alpha)\{f\}$.

Let $\succsim$ be a given binary relation on $\mathcal{M}$. Define a binary relation $\succsim^*$ on $\mathcal{K}_0$ using the operator m by way of the following definition: For all $K, K' \in \mathcal{K}_0$,

$$K \succsim^* K' \ :\Leftrightarrow \ m(K) \succsim m(K') \ . \tag{1.18}$$

The following Lemma characterizes some properties that $\succsim^*$ can inherit from $\succsim$.

Lemma 1 *Let $\succsim$ be a binary relation on $\mathcal{M}$ and define $\succsim^*$ on $\mathcal{K}_0$ by way of (1.18). Then $\succsim$ satisfies WO, VA, WSI, MON, MC and UB axioms if and only if $\succsim^*$ satisfies the WO*, AA*, WSI*, ΣMON*, MC* and UB* properties. Moreover, $\succsim$ satisfies SI if and only if $\succsim^*$ satisfies SI*.*

Proof. Let $\succsim$ be a binary relation on $\mathcal{M}$ and let $\succsim^*$ be defined as in (1.18).

Suppose $\succsim$ satisfies the WO, VA, WSI, MON, MC, and UB axioms, then we show that $\succsim^*$ satsfies WO*, AA*, WSI*, ΣMON*, MC*, and UB*.

(WO*): Let $K, K' \in \mathcal{K}_0$, then by WO, either $m(K) \succsim m(K')$ or $m(K') \succsim m(K)$, hence $\succsim^*$ is complete. Let $K, K', K'' \in \mathcal{K}_0$, and suppose that $K \succsim^* K'$ and $K' \succsim^* K''$. Then $m(K) \succsim m(K')$ and $m(K') \succsim m(K'')$, and so by WO it follows that $m(K) \succsim m(K'')$ so that $\succsim^*$ is transitive. Hence $\succsim^*$ is a weak order.

(AA*): Let $K, K' \in \mathcal{K}_0$ and $\alpha \in [0,1]$. By MON and Remark 2 it follows that $\alpha m(K) + (1-\alpha)m(K') \succsim m(\alpha K + (1-\alpha)K')$. Now if $K \sim^* K'$ then $m(K) \sim m(K')$ and by VA $m(K) \succsim \alpha m(K) + (1-\alpha)m(K') \succsim m(\alpha K + (1-\alpha)K')$, and so $K \succsim^* \alpha K + (1-\alpha)K'$. Hence $\succsim^*$ satisfies AA*.

(WSI*): Let $K, K' \in \mathcal{K}_0$, $K_f, K_g \in \mathcal{K}_S$ and $\alpha \in [0,1]$. Then by WSI and Remark 2,

$$\begin{aligned} m(\alpha K + (1-\alpha)K_f) &= \alpha m(K) + (1-\alpha)\{f\} \\ &\succsim \alpha m(K') + (1-\alpha)\{f\} \\ &= m(\alpha K' + (1-\alpha)K_f) \\ \Leftrightarrow \quad m(\alpha K + (1-\alpha)K_g) &= \alpha m(K) + (1-\alpha)\{g\} \\ &\succsim \alpha m(K') + (1-\alpha)\{g\} \\ &= m(\alpha K' + (1-\alpha)K_g)\,, \end{aligned}$$

and so $\alpha K + (1-\alpha)K_f \succsim^* \alpha K' + (1-\alpha)K_f$ iff $\alpha K + (1-\alpha)K_g \succsim^* \alpha K' + (1-\alpha)K_g$. Hence, $\succsim^*$ satisfies WSI*.

(ΣMON*): Let $K, K' \in \mathcal{K}_0$, and suppose that for all $\sigma \in \Sigma$ there exists $\sigma' \in \Sigma$ such that $K(s,\sigma') \succsim^* K'(s,\sigma)$ for all $s \in \mathcal{S}$. Let $g \in m(K')$. Then there exists some $\sigma \in \Sigma$ such that $g = K'(.,\sigma)$ and so there exists some $\sigma' \in \Sigma$ such that $K(s,\sigma') \succsim^* K'(s,\sigma)$ for all $s \in \mathcal{S}$. Let $f = K(s,\sigma')$. Then $f \in m(K)$ such that $f(s) \succsim g(s)$ for all $s \in \mathcal{S}$. By MON it follows that $m(K) \succsim m(K')$ and so $\succsim^*$ satisfies ΣMON*. Moreover, since ΣMON* implies MON*, $\succsim^*$ also satisfies MON*.

(MC*): Let $K_f, K_g \in \mathcal{K}_S$ and $K \in \mathcal{K}_0$. Then by Remark 2,

$$\left\{\alpha \in [0,1] \middle| \alpha K_f + (1-\alpha)K_g \succsim^* K\right\} = \left\{\alpha \in [0,1] \middle| m(\alpha K_f + (1-\alpha)K_g) \succsim m(K)\right\} \tag{1.19}$$

$$= \left\{\alpha \in [0,1] \middle| \alpha\{f\} + (1-\alpha)\{g\} \succsim m(K)\right\} \tag{1.20}$$

and

$$\left\{\alpha \in [0,1] \middle| K \succsim^* \alpha K_f + (1-\alpha)K_g\right\} = \left\{\alpha \in [0,1] \middle| m(K) \succsim m(\alpha K_f + (1-\alpha)K_g)\right\} \tag{1.21}$$

$$= \left\{\alpha \in [0,1] \middle| m(K) \succsim \alpha\{f\} + (1-\alpha)\{g\}\right\} \tag{1.22}$$

Since $\succsim$ satisfies MC all of the above sets are closed, and it follows that $\succsim^*$ satisfies MC*.

(UB*): By UB, there exists $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$ with $\{\mathbb{p}\} \succ \{\mathbb{q}\}$ such that for all $\alpha \in (0,1)$ there exists $\mathbb{r} \in \Delta(\mathcal{Z})$ satisfying either $\{\mathbb{q}\} \succ \alpha\{\mathbb{r}\} + (1-\alpha)\{\mathbb{p}\}$ or $\alpha\{\mathbb{r}\} + (1-\alpha)\{\mathbb{q}\} \succ \{\mathbb{p}\}$. By the natural identification of $\mathcal{K}_c \subset \mathcal{K}_0$ with $\Delta(\mathcal{Z})$, $\succsim^*$ therefore satisfies UB* as $m(K_{\mathbb{p}}) = \{\mathbb{p}\}$ by Remark 2.

Finally, suppose $\succsim$ satisfies SI.

(SI*): Let $K, K' \in \mathcal{K}_0$, $f \in \mathcal{F}$ and $\alpha \in (0,1)$. Then by SI and Remark 2, $K \succsim^* K'$ iff $m(K) \succsim m(K')$ iff $\alpha m(K) + (1-\alpha)\{f\} \succsim \alpha m(K') + (1-\alpha)\{f\}$ iff $m(\alpha K + (1-\alpha)K_f) \succsim m(\alpha K' + (1-\alpha)K_f)$ iff $\alpha K + (1-\alpha)K_f \succsim^* \alpha K' + (1-\alpha)K_f$. Hence, $\succsim^*$ satisfies SI*.

To show the converse, let $\succsim^*$ satisfy WO*, AA*, WSI*, ΣMON*, MC*, and UB*.

(WO): Let $M, N \in \mathcal{M}$. By WO* and Remark 2, $K_M \succsim^* K_N$ or $K_N \succsim^* K_M$, and so $\succsim$ is complete. Let $L, M, N \in \mathcal{M}$ such that $L \succsim M$ and $M \succsim N$. Then $K_L \succsim^* K_M$ and $K_M \succsim^* K_N$, and so by WO*, $K_L \succsim^* K_N$. Hence, $\succsim$ is a weak order.

(VA): Let $M = \{f_1, ..., f_{T_M}\}, N = \{g_1, ..., g_{T_N}\} \in \mathcal{M}$ and $\alpha \in [0,1]$. Let $K \in \mathcal{K}_0$ be defined as follows:

$$K(.,\sigma_j) = \begin{cases} f_1 & \text{if } j \in \{nT_N + 1 \mid n \in \mathbb{N}\} \\ f_2 & \text{if } j \in \{nT_N + 2 \mid n \in \mathbb{N}\} \\ \vdots & \\ f_{T_M} & \text{if } j \in \{nT_N + T_M \mid n \in \mathbb{N}\} \end{cases} .$$

Let $K' \in \mathcal{K}_0$ be defined as follows:

$$K'(.,\sigma_j) = \begin{cases} g_1 & \text{if } j \in \{nT_N T_M + k \mid t \in \mathbb{N} \text{ and } k \in \{1,...,T_M\}\} \\ g_2 & \text{if } j \in \{T_M + nT_N T_M + k \mid n \in \mathbb{N} \text{ and } k \in \{1,...,T_M\}\} \\ g_3 & \text{if } j \in \{2T_M + nT_N T_M + k \mid n \in \mathbb{N} \text{ and } k \in \{1,...,T_M\}\} \\ \vdots & \\ g_{T_N} & \text{if } j \in \{(T_N - 1)T_M + nT_N T_M + k \mid n \in \mathbb{N} \text{ and } k \in \{1,...,T_M\}\} \end{cases} .$$

Then $m(K) = M$, $m(K') = N$ and $m(\alpha K + (1-\alpha)K') = \alpha M + (1-\alpha)N$. Hence, if $M \sim N$ then $K \sim^* K'$, and so by AA*, $K \succsim^* \alpha K + (1-\alpha)K'$. Hence, $m(K) \succsim m(\alpha K + (1-\alpha)K')$ and therefore $M \succsim \alpha M + (1-\alpha)N$. Hence, $\succsim$ satisfies VA.

(WSI): Let $M, N \in \mathcal{M}$, $f, g \in \mathcal{F}$ and $\alpha \in (0,1)$. By WSI*,

$$\begin{aligned} & \alpha K_M + (1-\alpha)K_f \quad \succsim^* \quad \alpha K_N + (1-\alpha)K_f \\ \Leftrightarrow \quad & \alpha K_M + (1-\alpha)K_g \quad \succsim^* \quad \alpha K_N + (1-\alpha)K_g , \end{aligned}$$

and so by Remark 2, $\succsim$ satisfies WSI.

(MON): Let $M, N \in \mathcal{M}$ and suppose that for all $g \in N$ there exists $f \in M$ such that $\{f(s)\} \succsim \{g(s)\}$ for all $s \in \mathcal{S}$. Then for all $\sigma \in \Sigma$, there exists some $\sigma' \in \Sigma$ such that $K_M(.,\sigma') \succsim^* K_N(.,\sigma)$. By ΣMON*, $K_M \succsim^* K_N$ and so $M \succsim N$ and $\succsim$ satisfies MON.

(MC): Let $f, g \in \mathcal{F}$ and $M \in \mathcal{M}$. By Remark 2,

$$\begin{aligned}\left\{\alpha \in [0,1] \,\middle|\, \alpha\{f\} + (1-\alpha)\{g\} \succsim M\right\} &= \left\{\alpha \in [0,1] \,\middle|\, K_{\alpha f+(1-\alpha)g} \succsim^* K_M\right\} &(1.23)\\ &= \left\{\alpha \in [0,1] \,\middle|\, \alpha K_f + (1-\alpha)K_g \succsim^* K_M\right\} &(1.24)\end{aligned}$$

and

$$\begin{aligned}\left\{\alpha \in [0,1] \,\middle|\, M \succsim \alpha\{f\} + (1-\alpha)\{g\}\right\} &= \left\{\alpha \in [0,1] \,\middle|\, K_M \succsim^* K_{\alpha f+(1-\alpha)g}\right\} &(1.25)\\ &= \left\{\alpha \in [0,1] \,\middle|\, K_M \succsim^* \alpha K_f + (1-\alpha)K_g\right\} &(1.26)\end{aligned}$$

By MC*, all of the above sets are closed, and so $\succsim$ satisfies MC.

(UB):. By UB* there exists $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$ with $K_{\mathbb{p}} \succ^* K_{\mathbb{q}}$ such that for all $\alpha \in (0,1)$ there exists $\mathbb{r} \in \Delta(\mathcal{Z})$ satisfying either $K_{\mathbb{q}} \succ^* \alpha K_{\mathbb{r}} + (1-\alpha)K_{\mathbb{p}}$ or $\alpha K_{\mathbb{r}} + (1-\alpha)K_{\mathbb{q}} \succ K_{\mathbb{p}}$. By the natural identification of $\mathcal{M}_C \subset \mathcal{M}$ with $\Delta(\mathcal{Z})$, $\succsim$ therefore satisfies UB because $m(K_{\mathbb{p}}) = \{\mathbb{p}\}$ for all $\mathbb{p} \in \Delta(\mathcal{Z})$.

Finally, suppose $\succsim^*$ satisfies SI*.

(SI): Let $M, N \in \mathcal{M}$, $f \in \mathcal{F}$ and $\alpha \in (0,1)$. Then by SI*, $M \succsim N$ iff $K_M \succsim^* K_N$ iff $\alpha K_M + (1-\alpha)K_f \succsim^* \alpha K_N + (1-\alpha)K_f$, and so by Remark 2 it follows that $M \succsim N$ iff $\alpha M + (1-\alpha)f \succsim \alpha N + (1-\alpha)f$. Hence, $\succsim$ satisfies SI. ■

A representation for the binary relation on superacts

For any function $u : \mathcal{Z} \to \mathbb{R}$, let $cl(u)$ denote the convex closure of $u(\mathcal{Z}) := \{u \in \mathbb{R} | u = u(z) \text{ for some } z \in \mathcal{Z}\}$. For any $K \in \mathcal{K}_0$, we use the notation $Eu \circ K$ to denote $(s, \sigma) \mapsto Eu[K(s, \sigma)]$. Let $Eu \circ \mathcal{K}_0 = \{Eu \circ K \,|\, K \in \mathcal{K}_0\}$. Maccheroni et al. (2006, Lemma 28) establish the following Remark.

Remark 3 *Let $u : \mathcal{Z} \to \mathbb{R}$, then $Eu \circ \mathcal{K}_0 = B_0(\mathcal{S} \times \Sigma, cl(u))$.*

The following Lemma gives a first representation result for the binary relation $\succsim^*$ on superacts.

Lemma 2 *Let $\succsim^*$ be a binary relation on $\mathcal{K}_0$. The following two statements are equivalent:*

(i) $\succsim^*$ satisfies WO, AA*, WSI*, ΣMON*, MC*, and UB*.*

(ii) There exists an unbounded function $u : \mathcal{Z} \to \mathbb{R}$ and a normalized, Σ-monotonic, singleton-additive, convex niveloid $I : B_0(\mathcal{S} \times \Sigma, cl(u)) \to \mathbb{R}$, such that for all $K, K' \in \mathcal{K}_0$:

$$K \succsim^* K' \;\Leftrightarrow\; I(Eu \circ K) \geq I(Eu \circ K')\ . \tag{1.27}$$

Moreover, there exists a unique $\pi \in \Delta(\mathcal{S})$ such that $I(Eu \circ K_f) = \int_{\mathcal{S}} Eu[f(s)]\, d\pi(s)$ for all $f \in \mathcal{F}$, and u is unique up to positive affine transformations.

Proof. To show (i) implies (ii), let $\succsim^*$ satisfy WO*, AA*, WSI*, ΣMON*, MC*, and UB*. We proceed in several steps.

Step 1: Existence of $u : \mathcal{Z} \to \mathbb{R}$ and a normalized, Σ-monotonic niveloid I representing $\succsim^*$.

For any $K \in \mathcal{K}_0$, let $\mathbb{b}_K$ be the $\succsim^*$-maximal element of the range of K, and $\mathbb{w}_K$ be the $\succsim^*$-minimal element of K. By ΣMON*, $\mathbb{b}_K \succsim^* K \succsim^* \mathbb{w}_K$. By MC*, the sets $\{\alpha \,|\, \alpha\mathbb{b}_K + (1-\alpha)\mathbb{w}_K \succ^* K\}$ and $\{\alpha \,|\, K \succ^* \alpha\mathbb{b}_K + (1-\alpha)\mathbb{w}_K\}$ are open, and since $[0,1]$ is a connected set there exists an $\alpha_K \in [0,1]$ such that $\alpha_K\mathbb{b}_K + (1-\alpha_K)\mathbb{w}_K \sim^* K$. Hence, every $K \in \mathcal{K}_0$ has a constant-singleton equivalent $\mathbb{p}_K \in \mathcal{K}_c$, such that $K \sim^* \mathbb{p}_K$. It then follows exactly as in the proof of Lemma 28 in Maccheroni et al.

(2006), that there exists a non-constant function $u : \mathcal{Z} \to \mathbb{R}$ and a normalized niveloid $I : B_0(\mathcal{S} \times \Sigma, cl(u)) \to \mathbb{R}$ such that for all $K, K' \in \mathcal{K}_0$,

$$K \succsim^* K' \;\Leftrightarrow\; I(Eu \circ K) \geq I(Eu \circ K') \,. \tag{1.28}$$

Moreover, the function u is unique up to positive affine transformations, and by Lemma 29 in Maccheroni et al. (2006) u is unbounded. In addition, it is easily verified that ΣMON* implies that I is Σ-monotonic.

Step 2: Convexity of I.

Let $\alpha \in [0, 1]$ and $\varphi, \psi \in B_0(\mathcal{S} \times \Sigma, cl(u))$ such that $I(\varphi) = I(\psi)$. Then there exist $K, K' \in \mathcal{K}$ such that $Eu \circ K = \varphi$, $Eu \circ K' = \psi$ and so $K \sim^* K'$. By AA*, $K \succsim^* \alpha K + (1 - \alpha)K'$ and so

$$\begin{aligned} I(\varphi) &= I(Eu \circ K) \\ &\geq I(Eu \circ (\alpha K + (1 - \alpha)K')) \\ &= I(\alpha Eu \circ K + (1 - \alpha)Eu \circ K') \\ &= I(\alpha\varphi + (1 - \alpha)\psi) \end{aligned} \tag{1.29}$$

Let $\bar{I} : B_0(\mathcal{S} \times \Sigma, -cl(u)) \to \mathbb{R}$ be defined by $\bar{I}(\varphi) = -I(-\varphi)$ for all $\varphi \in B_0(\mathcal{S} \times \Sigma, -cl(u))$. By Remark 1, $\bar{I}$ is a normalized niveloid. Also, it is easy to verify that for any $\varphi, \psi \in B_0(\mathcal{S} \times \Sigma, -cl(u))$ such that $\bar{I}(\varphi) = \bar{I}(\psi)$ and $\alpha \in [0, 1]$, $\bar{I}(\alpha\varphi + (1-\alpha)\psi) \geq \bar{I}(\varphi)$ by (1.29). Thus Lemma 25 in Maccheroni et al. (2006) guarantees that $\bar{I}$ is concave, and by Remark 1 I must be convex.

Step 3: Existence of $\pi \in \Delta(\mathcal{S})$ such that $I(Eu \circ K_f) = \int_{\mathcal{S}} Eu[f(s)]d\pi(s)$ for all $f \in \mathcal{F}$.

Let $\succsim'$ be a binary relation on $\mathcal{F}$ defined by:

$$f \succsim' g \text{ if } K_f \succsim^* K_g \,, \tag{1.30}$$

and associate with it the asymmetric part $\succ'$ and symmetric part $\sim'$ in the usual sense. It is easy to verify that $\succsim'$ satisfies counterparts of WO*, WSI*, MC* and UB* (denoted WO′, WSI′, MC′ and UB′), and that it satisfies the property MON′: If $f(s) \succsim' g(s)$ for all $s \in \mathcal{S}$, then $f \succsim' g$.

We now show that $\succsim'$ also satisfies I′: For all $f, g \in \mathcal{F}$, $f \succsim' g$ iff $\alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h$ for all $h \in \mathcal{F}$ and $\alpha \in (0,1)$.

Let $f, g \in \mathcal{F}$ such that $f \sim' g$. Suppose for contradiction, without loss of generality, that there exists some $h \in \mathcal{F}$ such that $(1/2)f + (1/2)h \succ' (1/2)g + (1/2)h$. Then, by WSI′, $f \succ' (1/2)g + (1/2)f$ and similarly $(1/2)g + (1/2)f \succ' g$, implying that $f \succ' g$, a contradiction. Therefore, for any $h \in \mathcal{F}$, we must have $(1/2)f + (1/2)h \sim' (1/2)g + (1/2)h$. Thus, together with properties WO′, MON′, MC′ and UB′, all three postulates in Herstein and Milnor (1953) are satisfied, and so by their Theorems 2 and 3, $\succsim'$ satisfies I′. It then follows from a standard result (see, e.g., Fishburn, 1970; Kreps, 1988, pp. 176–177, pp.99–111) that there exist a unique probability measure $\pi \in \Delta(\mathcal{S})$ and a non-constant function $v : \mathcal{Z} \to \mathbb{R}$, unique up to positive affine transformations, such that for all $f, g \in \mathcal{F}$,

$$f \succsim' g \;\Leftrightarrow\; \int_{\mathcal{S}} Ev[f(s)]\, d\pi(s) \;\geq\; \int_{\mathcal{S}} Ev[g(s)]\, d\pi(s) \,. \tag{1.31}$$

Since I is normalized, by (1.28), (1.30), (1.31) and the uniqueness properties of v and u it is without loss of generality to assume $v = u$ (hence, v is also unbounded). Thus for all $f, g \in \mathcal{K}_S$,

$$f \succsim' g \;\Leftrightarrow\; \int_{\mathcal{S}} Eu[f(s)]\, d\pi(s) \;\geq\; \int_{\mathcal{S}} Eu[g(s)]\, d\pi(s) \,. \tag{1.32}$$

Moreover, by (1.28), (1.30) and (1.32),

$$I\left(Eu \circ K_f\right) = I\left(Eu \circ K_{\mathbb{p}_f}\right) = Eu\left[\mathbb{p}_f\right] = \int_{\mathcal{S}} Eu[f(s)]\, d\pi(s)\ . \tag{1.33}$$

Step 4: Singleton-additivity of I.

To complete the proof that (i) implies (ii) we now show that WSI* also implies that I is singleton-additive. Since u is unbounded and unique only up to positive affine transformations, assume w.l.o.g that $0 \in$ int $(u(\mathcal{Z}))$. Now let $K \in \mathcal{K}_0$, $f \in \mathcal{F}$, and $\alpha \in [0,1]$, and note that $Eu[\alpha \mathbb{b}_K + (1-\alpha)K_f(s,\sigma)] \geq Eu[\alpha K(s,\sigma) + (1-\alpha)K_f(s,\sigma)] \geq Eu[\alpha \mathbb{w}_K + (1-\alpha)K_f(s,\sigma)]$ for all $(s,\sigma) \in \mathcal{S} \times \Sigma$. By monotonicity of I, it follows that

$$\begin{aligned} I\left(Eu \circ \left[\alpha K_{\mathbb{b}_K} + (1-\alpha)K_f\right]\right) &\geq I\left(Eu \circ \left[\alpha K + (1-\alpha)K_f\right]\right) \\ &\geq I\left(Eu \circ \left[\alpha K_{\mathbb{w}_K} + (1-\alpha)K_f\right]\right)\ , \end{aligned} \tag{1.34}$$

and so by (1.33),

$$\begin{aligned} \int_{\mathcal{S}} Eu\left[\alpha \mathbb{b}_K + (1-\alpha)f(s)\right]\, d\pi(s) &\geq I\left(Eu \circ \left[\alpha K + (1-\alpha)K_f\right]\right) \\ &\geq \int_{\mathcal{S}} Eu\left[\alpha \mathbb{w}_K + (1-\alpha)f(s)\right]\, d\pi(s)\ . \end{aligned} \tag{1.35}$$

Hence, there exists some $\beta \in [0,1]$ such that

$$\begin{aligned} I\left(Eu \circ \left[\alpha K + (1-\alpha)K_f\right]\right) &= \int_{\mathcal{S}} \left[\alpha\, Eu\left[\bar{\mathbb{p}}\right] + (1-\alpha)\, Eu[f(s)]\right] d\pi(s) \\ &= \alpha Eu\left[\bar{\mathbb{p}}\right] + (1-\alpha) I\left(Eu \circ K_f\right)\ , \end{aligned} \tag{1.36}$$

where $\bar{\mathbb{p}} := \beta \mathbb{b}_K + (1-\beta)\mathbb{w}_K$. By (1.28), (1.33), and (1.36), $\alpha K_{\bar{\mathbb{p}}} + (1-\alpha)K_f \sim^* \alpha K + (1-\alpha)K_f$. Let $\mathbb{p}_0 \in \Delta(\mathcal{Z})$ such that $Eu[\mathbb{p}_0] = 0$, then by WSI*, $\alpha K_{\bar{\mathbb{p}}} + (1-\alpha)K_{\mathbb{p}_0} \sim^*$

$\alpha K + (1-\alpha) K_{\mathbb{p}_0}$, and hence

$$\begin{aligned}
\alpha Eu[\bar{\mathbb{p}}] &= \alpha Eu(\bar{\mathbb{p}}) + (1-\alpha) Eu(\mathbb{p}_0) \\
&= Eu\left(\alpha \bar{\mathbb{p}} + (1-\alpha)\mathbb{p}_0\right) \\
&= I\Big(Eu \circ \left(\alpha K_{\bar{\mathbb{p}}} + (1-\alpha) K_{\mathbb{p}_0}\right)\Big) \\
&= I\Big(Eu \circ \left(\alpha K + (1-\alpha) K_{\mathbb{p}_0}\right)\Big) \\
&= I\Big(\alpha\, Eu \circ K + (1-\alpha)\, Eu \circ K_{\mathbb{p}_0}\Big) \\
&= I\Big(\alpha\, Eu \circ K + 0\Big) = I\Big(\alpha\, Eu \circ K\Big) \qquad (1.37)
\end{aligned}$$

Hence,

$$I(Eu \circ \Big[\alpha K + (1-\alpha) K_f\Big]) = I(\alpha\, Eu \circ K) + (1-\alpha) I(Eu \circ K_f)\ , \qquad (1.38)$$

and so I is singleton-additive.

Steps 1 - 4 complete the proof that (i) implies (ii), and establish the existence of $\pi \in \Delta(\mathcal{S})$ by Step 3.

To prove that (ii) implies (i), let $u : \mathcal{Z} \to \mathbb{R}$ be an unbounded function, and $I : B_0(\mathcal{S} \times \Sigma, cl(u)) \to \mathbb{R}$ be a normalized, Σ-monotone, singleton-additive, convex niveloid that represents $\succsim^*$ in the sense of (1.28). Lemmas 28 and 29 in Maccheroni et al. (2006) shows that $\succsim^*$ satisfies WO*, MC*, UB*, ΣMON*, and WSCI*. Property ΣMON* follows directly from Σ-monotonicity of I, and property AA* follows directly from convexity of I. It therefore remains to show that $\succsim^*$ satisfies WSI*. Let $K, K' \in \mathcal{K}_0$, $f, g \in \mathcal{F}$ and $\alpha \in (0,1)$ such that

$\alpha K + (1-\alpha)K_f \succsim^* \alpha K' + (1-\alpha)K_f$. Then by (1.28),

$$\begin{aligned}
& I\Big(Eu \circ \big(\alpha K + (1-\alpha)K_f\big)\Big) && \geq && I\Big(Eu \circ \big(\alpha K' + (1-\alpha)K_f\big)\Big) \\
\Leftrightarrow \quad & I\Big(\alpha(Eu \circ K)\Big) + (1-\alpha)I\Big(Eu \circ K_f\Big) && \geq && I\Big(\alpha(Eu \circ K')\Big) + (1-\alpha)I\Big(Eu \circ K_f\Big) \\
\Leftrightarrow \quad & I\Big(\alpha Eu \circ K\Big) + (1-\alpha)I\Big(Eu \circ K_g\Big) && \geq && I\Big(\alpha Eu \circ K'\Big) + (1-\alpha)I\Big(Eu \circ K_g\Big) \\
\Leftrightarrow \quad & I\Big(Eu \circ (\alpha K + (1-\alpha)K_g)\Big) && \geq && I\Big(Eu \circ \big(\alpha K' + (1-\alpha)K_g\big)\Big) \\
\Leftrightarrow \quad & \alpha K + (1-\alpha)K_g && \succsim^* && \alpha K' + (1-\alpha)K_g
\end{aligned} \tag{1.39}$$

showing that $\succsim^*$ satisfies WSI*, and also completing the proof that (ii) implies (i). Since, (i) and (ii) are equivalent, Step 3 in the proof that (i) implies (ii) establishes the existence of $\pi \in \Delta(\mathcal{S})$ such that $I(Eu \circ K_f) = \int_{\mathcal{S}} Eu[f(s)]d\pi(s)$ for all $f \in \mathcal{F}$. ■

The following Lemma is a central step in identifying the information cost function in Theorem 1. Here, $\Delta^*(\mathcal{S} \times \Sigma)$ refers to the set of all *finitely additive* probability measures on $2^{\mathcal{S}\times\Sigma}$, a set which clearly contains $\Delta(\mathcal{S} \times \Sigma)$.

Lemma 3 *Let $u : \mathcal{Z} \to \mathbb{R}$ be an unbounded function with $0 \in int\ cl(u)$, and I be a real-valued functional on $B_0(\mathcal{S} \times \Sigma, cl(u))$. Then (in view of Remark 3) the following statements are equivalent.*

(i) The functional I is a normalized, Σ-monotonic, singleton-additive, convex niveloid.

(ii) For all $K \in \mathcal{K}_0$, the functional I satisfies

$$I(Eu \circ K) = \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \left(\int_{\Sigma} \left[\max_{f \in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\, p(s|\sigma) \right] d\, p_{\Sigma}(\sigma) \; - \; c^*(p) \right), \tag{1.40}$$

where $c^* : \Delta^*(\mathcal{S} \times \Sigma) \to [0, \infty]$, defined by

$$c^*(p) = \sup_{K \in \mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, p \; - \; I(Eu \circ K) \right) \quad \forall p \in \Delta^*(\mathcal{S} \times \Sigma)\,, \tag{1.41}$$

is the unique grounded, convex, weak lower-semicontinuous function satisfying (1.40) and*

(a) $c^*(p) = \infty$ *for all* $p \in \Delta^*(\mathcal{S} \times \Sigma)$ *such that* $p_{\mathcal{S}} \neq \pi$ *for some unique* $\pi \in \Delta(\mathcal{S})$.

Moreover, in this case $c^*(p)$ *also satisfies the following three properties:*

(b) $c^*(p) \geq c^*(q)$ *if* $p, q \in \Delta^*(\mathcal{S} \times \Sigma)$ *such that* $p \succsim q$.

(c) $c^*(p) = 0$ *for all* $p \in \bar{\Delta}(\mathcal{S} \times \Sigma)$ *such that* $p_{\mathcal{S}} = \pi$.

(d) $c^*(p) = \infty$ *for all* $p \in \Delta^*(\mathcal{S} \times \Sigma) \backslash \Delta(\mathcal{S} \times \Sigma)$.

Proof. Let $\bar{I} : B_0(\mathcal{S} \times \Sigma, -cl(u)) \to \mathbb{R}$ be defined by $\bar{I}(Eu \circ K) := -I(-(Eu \circ K))$ for all $(Eu \circ K) \in B_0(\mathcal{S} \times \Sigma, -cl(u))$. By Remark 1, $\bar{I}$ is a concave and normalized niveloid. Thus, by Lemma 26 in Maccheroni et al. (2006), $\bar{I}$ satisfies

$$\bar{I}(Eu \circ K) = \min_{p \in \Delta^*(\mathcal{S} \times \Sigma)} \left(\int_{\mathcal{S} \times \Sigma} (Eu \circ K)\, d\, p \;+\; \bar{c}(p) \right) \quad \forall (Eu \circ K) \in B_0(\mathcal{S} \times \Sigma, -\, cl(u)) \,, \tag{1.42}$$

where $\bar{c}(p)$ is a non-negative, grounded, convex, and weak* lower semi-continuous function defined on $\Delta^*(\mathcal{S} \times \Sigma)$. Since u is an unbounded function, then again by Lemma 26 in Maccheroni et al. (2006),

$$\begin{aligned} \bar{c}^*(p) &= -\inf_{(Eu \circ K') \in B_0(\mathcal{S} \times \Sigma), -cl(u)} \left(\int_{\mathcal{S} \times \Sigma} (Eu \circ K')\, d\, p \;-\; \bar{I}(Eu \circ K') \right) \\ &= \sup_{K \in \mathcal{K}_0} \Big(\bar{I}(\,-\,(Eu \;\circ\; K)) \;-\; \int_{\mathcal{S} \times \Sigma} -\,(Eu \circ K)\, d\, p \Big) \end{aligned} \tag{1.43}$$

is the unique non-negative, grounded, convex and weak* lower semi-continuous function satisfying (1.42).

Thus, by using the definition of $\bar{I}$ and (1.42), for all $Eu \circ K \in B_0(\mathcal{S} \times \Sigma, cl(u))$,

$$\begin{aligned} I(Eu \circ K) &= -\bar{I}(-(Eu \circ K)) \\ &= -\min_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \left(\int_{\mathcal{S}\times\Sigma} -(Eu \circ K)\, d\,p \,+\, \bar{c}^*(p) \right) \\ &= \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\,p \,-\, c^*(p) \right) , \end{aligned} \tag{1.44}$$

where, by the definition of $\bar{I}$ and (1.43),

$$\begin{aligned} c^*(p) &= \sup_{(Eu\circ K') \in B_0(\mathcal{S}\times\Sigma, cl(u))} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K')\, d\,p \,-\, I(Eu \circ K') \right) \\ &= \sup_{K \in \mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\,p \,-\, I(Eu \circ K) \right) \\ &= \sup_{K \in \mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\,p \,+\, \bar{I}(-(Eu \circ K)) \right) = \bar{c}^*(p) , \end{aligned} \tag{1.45}$$

is the unique non-negative, grounded, convex and weak* lower semi-continuous function defined on $\Delta^*(\mathcal{S} \times \Sigma)$ satisfying (1.44).

Now let $K \in \mathcal{K}_0$ and observe that, for all $p \in \Delta^*(\mathcal{S} \times \Sigma)$,

$$\begin{aligned} J(m(K), p) &:= \int_{\Sigma} \left(\max_{f \in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\,p(s|\sigma) \right) d\,p_{\Sigma}(\sigma) \\ &\geq \int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\,p . \end{aligned} \tag{1.46}$$

Thus by (1.44) and (1.46), we have

$$I(Eu \circ K) \leq \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} J(m(K), p) \,-\, c^*(p) . \tag{1.47}$$

Let $p^* \in \Delta^*(\mathcal{S} \times \Sigma)$ such that $p^* \in \arg\max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} J(m(K), p) - c^*(p)$ and define a function $K^* : \mathcal{S} \times \Sigma \to \mathbb{R}$ such that

$$K^*(., \sigma) \in \arg\max_{f \in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\,p(s|\sigma) \quad \forall \sigma \in \Sigma .$$

It is clear that $K^* \in \mathcal{K}_0$. By (1.44)-(1.46), $I(Eu \circ K^*) = J(m(K), p^*) - c^*(p^*)$, and so by (1.47), $I(Eu \circ K) \leq I(Eu \circ K^*)$. Moreover, by definition of K^*, $(Eu \circ K) \geq_{\Sigma} (Eu \circ K^*)$, and so by Σ-monotonicity of I, we must have $I(Eu \circ K) \geq I(Eu \circ K^*)$, implying that

$$I(Eu \circ K) = \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} J(m(K), p) - c^*(p)$$
$$= \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \int_\Sigma \left[\max_{f \in m(K)} \int_\mathcal{S} Eu[f(s)]\, d\, p(s|\sigma) \right] d\, p_\Sigma(\sigma) - c^*(p) \quad (1.48)$$

Next, we verify the properties (a) – (d) of the function c^*. Note that since $I : B_0(\mathcal{S}\times\Sigma, cl(u)) \to \mathbb{R}$ is a normalized niveloid which is Σ-monotone and singleton-additive, by Lemma 2 there exists a unique $\pi \in \Delta(\mathcal{S})$ such that $I(Eu \circ K_f) = \int_\mathcal{S} Eu[f(s)]\, d\, \pi(s)$ for all acts $f \in \mathcal{F}$.

First consider the case of u unbounded above and let $p \in \Delta^*(\mathcal{S} \times \Sigma)$ such that $p_S \neq \pi$. Then there exists $s \in \mathcal{S}$ such that $p_\mathcal{S}(s) > \pi(s)$. For such an $s \in \mathcal{S}$, let $\bar{\mathcal{K}}_S := \{K_f \in \mathcal{K}_S : Eu[f(s)] = a \text{ for all } s' \in \mathcal{S}\backslash\{s\}\}$ for some fixed $a \in cl(u)$. Then by (1.45) and the fact that $I(Eu \circ K_f) = \int_S Eu[f(s)] d\pi(s)$ for all $f \in \mathcal{F}$:

$$\begin{aligned} c^*(p) &= \sup_{K \in \mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, p - I(Eu \circ K) \right) \\ &\geq \sup_{K_f \in \bar{\mathcal{K}}_S} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, p - I(Eu \circ K_f) \right) \\ &= \sup_{K_f \in \bar{\mathcal{K}}_S} \left(\int_\mathcal{S} Eu[f(s)]\, d\, p_\mathcal{S}(s) - \int_\mathcal{S} Eu[f(s)]\, d\, \pi(s) \right) \\ &= (p_\mathcal{S}(s) - \pi(s)) \sup_{\mathbb{p} \in \Delta(\mathcal{Z})} (Eu[\mathbb{p}] - a) = \infty\,. \end{aligned} \quad (1.49)$$

A similar argument shows that when u is unbounded below, $c^*(p) = \infty$, and so we conclude that $c^*(p)$ satisfies property (a).

To show that c^* satisfies property (b), let $p, q \in \Delta(\mathcal{S} \times \Sigma)$ such that $p \succsim q$. Let $K \in \mathcal{K}_0$ and define $K^p \in \mathcal{K}_0$ such that for all $\sigma \in \Sigma$,

$$K^p(., \sigma) \in \arg\max_{f \in m(K)} \int_\mathcal{S} Eu[f(s)]\, d\, p(s|\sigma) \quad (1.50)$$

Then it is clear that

$$\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, q \;\leq\; \int_{\mathcal{S}\times\Sigma} (Eu \circ K^p)\, d\, p\,. \tag{1.51}$$

Moreover, by construction of the superact K^p, $K \geq_\Sigma K^p$ and so by Σ-monotonicity of I, $I(Eu \circ K) \geq I\,(Eu \circ K^p)$. Combining these two inequalities,

$$\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, q \;-\; I(Eu \circ K) \;\leq\; \int_{\mathcal{S}\times\Sigma} (Eu \circ K^p)\; d\, p \;-\; I\,(Eu \circ K^p)\;. \tag{1.52}$$

From the last inequality, it follows that

$$\begin{aligned} c^*(q) &= \sup_{K\in\mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, q \;-\; I(Eu \circ K) \right) \\ &\leq \sup_{K^p\in\mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K^p)\; d\, p \;-\; I\,(Eu \circ K^p) \right) \\ &\leq \sup_{K\in\mathcal{K}_0} \left(\int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\, p \;-\; I(Eu \circ K) \right) = c^*(p)\,, \end{aligned} \tag{1.53}$$

and hence c^* satisfies property (b).

To show that c^* satisfies property (c), let $p_\Sigma \in \Delta^*(\Sigma)$. Since I is Σ-monotonic, for any $K \in \mathcal{K}_0$,

$$\begin{aligned} & \int_{\mathcal{S}\times\Sigma} (Eu \,\circ\, K)\, d\; (\pi \times p_\Sigma) \;-\; I(Eu \,\circ\, K) \\ \leq\;& \int_{\Sigma} \left(\max_{f\in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\, \pi(s) \right) d\, p_\Sigma \;-\; I(Eu \,\circ\, K) \\ =\;& \max_{f\in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\, \pi(s) \;-\; I(Eu \,\circ\, K) \\ =\;& I(Eu \,\circ K_f) \;-\; I(Eu \,\circ\, K) \leq 0 \end{aligned} \tag{1.54}$$

and in particular,

$$I(Eu \,\circ\, K_f) \;-\; I(Eu \,\circ\, K) = 0. \tag{1.55}$$

if $K = K_f$ for some $f \in \mathcal{F}$. Hence, by (1.45) and non-negativity of c^*, we have $c^*(\pi \times p_\Sigma) = 0$, and so property (c) is satisfied.

Finally, we verify that c^* also satisfies property (d). To see this, note that for any sequence of sets $E_1, E_2, ... \in 2^{\mathcal{S}\times\Sigma}$ such that $E_1 \supset E_2 \supset ...$, and $\bigcap_{n\geq 1} E_n = \emptyset$, there exists some $N \in \mathbb{N}$ such that $n \geq N$ implies $E_n = \emptyset$. Hence, by Proposition 29 and Theorem 11 in Maccheroni et al. (2004), and the unboundedness of u, $\{p \in \Delta^*(\mathcal{S} \times \Sigma) | c^*(p) \leq t\}$ is a weak* compact subset of $\Delta(\mathcal{S} \times \Sigma)$ for all $t \geq 0$. Therefore c^* satisfies property (d), and this concludes the proof that (i) implies (ii).

For the converse, let I be a real-valued functional on $B_0(\mathcal{S} \times \Sigma, cl(u))$, defined for all $K \in \mathcal{K}_0$ by

$$I(Eu \circ K) = \max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} \left(\int_\Sigma \left[\max_{f\in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\, p(s|\sigma) \right] d\, p_\Sigma(\sigma) - c(p) \right) \tag{1.56}$$

for some grounded, convex, weak* lower semi-continuous function $c : \Delta^*(\mathcal{S} \times \Sigma) \to [0, \infty]$ satisfying property (a).

We first show that I is a niveloid. Note that since $c(.)$ is grounded, by (1.56), $I(Eu \circ K_{\mathbb{p}}) = Eu[\mathbb{p}]$ for any $\mathbb{p} \in \Delta(\mathcal{Z})$. That is, I is normalized. Let $K, K' \in \mathcal{K}_0$ such that $Eu \circ K \geq Eu \circ K'$. Then for any $p \in \Delta^*(\mathcal{S} \times \Sigma)$

$$J(m(K), p) \geq J(m(K'), p) \ , \tag{1.57}$$

where $J(m(K), p)$ is defined as in (1.46) and thus, by (1.56), $I(Eu \circ K) \geq I(Eu \circ K')$, showing that I is monotonic. Now let $K \in \mathcal{K}_0$, $\mathbb{p} \in \Delta(\mathcal{Z})$, and $\alpha \in [0, 1]$. Thus, we have $\alpha K + (1-\alpha)K_{\mathbb{p}} \in \mathcal{K}_0$ and, by (1.46),

$$
\begin{aligned}
I\Big(\alpha(Eu \circ K) + (1-\alpha)Eu[\mathbb{p}]\Big) &= I\Big(\alpha(Eu \circ K) + (1-\alpha)(Eu \circ K_{\mathbb{p}})\Big) \\
&= I\Big(Eu \circ (\alpha K + (1-\alpha)K_{\mathbb{p}})\Big) \\
&= \max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} \left(J(m(\alpha K + (1-\alpha)K_{\mathbb{p}}), p) - c(p)\right) \\
&= \max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} \left(\alpha J(m(K), p) + (1-\alpha)Eu[\mathbb{p}] - c(p)\right) \\
&= \max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} \left(\alpha J(m(K), p) - c(p)\right) + (1-\alpha)Eu[\mathbb{p}] \\
&= I(\alpha\,(Eu \circ K)) + (1-\alpha)Eu[\mathbb{p}]\ . \qquad (1.58)
\end{aligned}
$$

Since $0 \in \text{int } cl(u)$ and I satisfies monotonicity and (1.58), by Lemma 25 in Maccheroni et al. (2006), I is a niveloid.

To show that I is convex, let $K, K' \in \mathcal{K}_0$ and $\alpha \in [0,1]$. Then by (1.56),

$$
\begin{aligned}
& I\Big(\alpha(Eu \circ K) + (1-\alpha)(Eu \circ K')\Big) \\
= {} & I\Big(Eu \circ (\alpha K + (1-\alpha)K')\Big) \\
= {} & \max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} \Big(J(m(\alpha K + (1-\alpha)K'), p) - c(p)\Big) \\
= {} & \max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} \Big(\alpha(J(m(K), p) - c(p)) + (1-\alpha)\big(J(m(K'), p) - c(p)\big)\Big) \\
\leq {} & \alpha\Big(\max_{p\in\Delta^*(\mathcal{S}\times\Sigma)} J(m(K), p) - c(p)\Big) + (1-\alpha)\Big(\max_{q\in\Delta^*(\mathcal{S}\times\Sigma)} J(m(K'), q) - c(q)\Big) \\
= {} & \alpha I(Eu \circ K) + (1-\alpha)I(Eu \circ K') \qquad (1.59)
\end{aligned}
$$

as desired.

Since I is a convex, normalized niveloid and u is unbounded, by the observations on (1.42)–(1.45) and Lemma 25 in Maccheroni et al. (2006) we deduce that c must be equal to the function c^* defined in (1.45), and conclude that c is the unique function satisfying (1.56).

Next, we verify that I is also singleton-additive. First, let $f \in \mathcal{F}$ and note that

$$\begin{aligned}
I(Eu \circ K_f) &= \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \Big(J\Big(m\Big(K_f\Big), p\Big) - c(p)\Big) \\
&= \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \Big(\int_{\mathcal{S}} Eu[f(s)]\, d\, p_{\mathcal{S}}(s) - c(p)\Big) \\
&= \int_{\mathcal{S}} Eu[f(s)]\, d\,\pi(s) \; - \min_{p \in \Delta^*(\mathcal{S}\times\Sigma)} c(p) \quad \text{(by property (a))} \\
&= \int_{\mathcal{S}} Eu[f(s)]\, d\,\pi(s) \quad \text{(by groundedness)}
\end{aligned} \tag{1.60}$$

Now let $K \in \mathcal{K}_0$, $f \in \mathcal{F}$ and $\alpha \in [0,1]$. By (1.46), (1.56), and (1.60),

$$\begin{aligned}
& I\Big(\alpha(Eu \circ K) + (1-\alpha)(Eu \circ K_f)\Big) \\
= \;& I\Big(Eu \circ (\alpha K + (1-\alpha)K_f)\Big) \\
= \;& \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \Big(J(m[\alpha K + (1-\alpha)K_f], p) - c(p)\Big) \\
= \;& \max_{p \in \Delta^*(\mathcal{S}\times\Sigma)} \Big(\alpha J(m[K], p) - c(p)\Big) + \int_{\mathcal{S}} Eu[f(s)]\, d\,\pi(s) \\
= \;& I(\alpha(Eu \circ K)) + (1-\alpha) I(Eu \circ K_f)
\end{aligned} \tag{1.61}$$

and so I is singleton-additive.

Finally, to show that I is Σ-monotone, let $K, K' \in \mathcal{K}_0$ such that $(Eu \circ K) \geq_{\Sigma} (Eu \circ K')$. Then for all $p \in \Delta^*(\mathcal{S}\times\Sigma)$,

$$\max_{f \in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\, p(s|\sigma'') \geq \max_{f \in m(K')} \int_{\mathcal{S}} Eu[f(s)]\, d\, p(s|\sigma'')\,, \tag{1.62}$$

for all $\sigma'' \in \Sigma$. Hence, for all $p \in \Delta^*(\mathcal{S}\times\Sigma)$

$$J(m(K), p) \geq J(m(K'), p)\,, \tag{1.63}$$

and so by (1.46) and (1.56), $I(Eu \circ K) \geq I(Eu \circ K')$, showing that I is Σ-monotone. This completes the proof. ■

1.7.3 Proofs of results in the paper

Proof of Theorem 1.

Let $\succsim$ be a given binary relation on $\mathcal{M}$ and define a binary relation $\succsim^*$ on $\mathcal{K}_0$ by way of (1.18).

(i) implies (ii): Suppose that $\succsim$ satisfies the WO, VA, WSI, MON, MC, and UB axioms. Then by Lemma 1, $\succsim^*$ satisfies WO*, AA*, WSI*, ΣMON*, MC* and UB*. Thus, by Lemma 2 there exists an unbounded function $u : \mathcal{Z} \to \mathbb{R}$, a prior $\pi \in \Delta(\mathcal{S})$ and a normalized, Σ-monotonic, singleton additive, convex niveloid $I : Eu \circ \mathcal{K}_0 \to \mathbb{R}$ such that for all $K, K' \in \mathcal{K}_0$

$$K \succsim^* K' \;\Leftrightarrow\; I(Eu \circ K) \geq I(Eu \circ K')\,, \tag{1.64}$$

and for all $f \in \mathcal{F}$, $I(Eu \circ K_f) = \int_{\mathcal{S}} Eu[f(s)]d\pi(s)$, and it is w.l.o.g to let $0 \in \text{int}\,(u(\mathcal{Z}))$.

By Lemma 3, there exists an information cost function $c^* : \Delta(\mathcal{S} \times \Sigma) \to [0, \infty]$ (relative to π), such that for all $K \in \mathcal{K}_0$,[14]

$$I(Eu \circ K) = \max_{p \in \Delta(\mathcal{S}\times\Sigma)} \left(\int_{\Sigma} \left[\max_{f \in m(K)} \int_{\mathcal{S}} Eu[f(s)]\, d\, p(s|\sigma) \right] d\, p_{\Sigma}(\sigma) \;-\; c^*(p) \right)\,. \tag{1.65}$$

Define $V(M) := I(Eu \circ K_M)$ for all $M \in \mathcal{M}$, then by (1.18) and (1.64), V is a representation of $\succsim$ on $\mathcal{M}$. Application of the law of iterated expectations gives the desired representation in Theorem 1, and the uniqueness properties of u, π and c^* follow from Lemma 1, Lemma 2 and Lemma 3.

(ii) implies (i): Let $u : \mathcal{Z} \to \mathbb{R}$ be an unbounded utility index with $0 \in \text{int}\,(u(\mathcal{Z}))$, and $c : \Delta(S \times \Sigma) \to [0, \infty]$ be an information cost function relative

[14] In particular, it is property (d) of the function $c^* : \Delta^*(\mathcal{S} \times \Sigma)$ in Lemma 3, that ensures that the cost function here can be defined on $\Delta(\mathcal{S} \times \Sigma)$.

to some $\pi \in \Delta(\mathcal{S})$, such that the function $V : \mathcal{M} \to \mathbb{R}$ defined for all $M \in \mathcal{M}$ as

$$V(M) = \max_{p \in \Delta(\mathcal{S} \times \Sigma)} \int_{\Sigma} \left[\max_{f \in M} \int_{\mathcal{S}} Eu[f(s)]\, d\, p(s|\sigma) \right] d\, p_{\Sigma}(\sigma) \; - \; c^{*}(p) \; , \tag{1.66}$$

represents the binary relation $\succsim$ on $\mathcal{M}$. Define the function $I : Eu \circ \mathcal{K}_0 \to \mathbb{R}$ as for all $K \in \mathcal{K}_0$, $I(Eu \circ K) = V(m(K))$. Then I satisfies all the premises of Lemma 3 (ii), and so by Lemma 3 I is a normalized, Σ-monotonic, singleton-additive and convex niveloid on $Eu \circ \mathcal{K}_0$. Moreover, since V represents $\succsim$, by definition of the binary relation $\succsim^{*}$, I represents $\succsim^{*}$ on $\mathcal{K}_0$. Thus, by Lemma 2, $\succsim^{*}$ must satisfy properties WO*, AA*, WSI*, ΣMON*, MC* and UB* on $\mathcal{K}_0$. This implies that $\succsim$ satisfies properties WO, VA, WSI, MON, MC and UB on $\mathcal{M}$ by Lemma 1. ■

Proof of Observation 1. Define for each $M \in \mathcal{M}$ an *oracle act* of M by, for all $s \in S$

$$O_M(s) = f(s) \text{ for some } f \in M \text{ such that } \{f(s)\} \succsim \{g(s)\} \text{ for all } g \in M. \tag{1.67}$$

The observation follows directly as (i) V represents $\succsim$, (ii) $\succsim$ satisfies MON, (iii) MON implies that for all $M \in \mathcal{M}$, $M \succsim \{f\}$ if $f \in M$, and (iv) MON implies that $\{O_M\} \succsim M$. ■

Proof of Proposition 1. For all $M \in \mathcal{M}$, the function V^{*} satisfies

$$V^{*}(M) = \max_{p \in \Delta(\mathcal{S} \times \Sigma)} \left(\int_{\mathcal{S}} \left(\int_{\Sigma} \left(\max_{f \in M} \int_{\mathcal{S}} Eu[f(s)] dp(s|\sigma) \right) dp(\sigma|s) \right) dp(s) - \tilde{c}(p) \right) , \tag{1.68}$$

where

$$\tilde{c}(p) = \begin{cases} \theta \mathrm{MI}(p) & \text{if } p_{\mathcal{S}} = \pi \\ \infty & \text{otherwise} \end{cases} . \tag{1.69}$$

Since mutual information is a non-negative, weak* lower-semicontinuous, convex function on $\Delta(\mathcal{S} \times \Sigma)$, $\tilde{c}$ is clearly non-negative, weak* lower-semicontinuous and convex. We now verify that $\tilde{c}$ also satisfies all three other properties of an information cost function. By definition of $\tilde{c}$, property (i) is satisfied and since $MI(p) = 0$ when $p \in \bar{\Delta}(\mathcal{S} \times \Sigma)$, $\tilde{c}$ also satisfies property (ii).

It remains to show that $\tilde{c}$ also satisfies property (iii). Let $p, q \in \Delta(\mathcal{S} \times \Sigma)$ such that $p \succsim q$. Then there must exist some Markov kernel $\mathbb{K}$ on $\Sigma \times 2^{\Sigma}$ satisfying for all $(s, \sigma') \in \Delta(\mathcal{S} \times \Sigma)$, $\quad q(s, \sigma') = \int_{\Sigma} p(s, \sigma)\mathbb{K}(\sigma, \sigma')d\sigma$. Hence q and p must satisfy,

$$\begin{aligned} q_S(s) &= \int_{\Sigma} q(s, \sigma')d\sigma' \\ &= \int_{\Sigma}\left(\int_{\Sigma} p(s, \sigma)\mathbb{K}(\sigma, \sigma')d\sigma\right)d\sigma' \\ &= \int_{\Sigma} p(s, \sigma)\left(\int_{\Sigma} \mathbb{K}(\sigma, \sigma')d\sigma'\right)d\sigma \\ &= \int_{\Sigma} p(s, \sigma)d\sigma = p_S(s) \end{aligned} \tag{1.70}$$

Thus if $p_S(s) \neq \pi$, then $\tilde{c}(p) = \tilde{c}(q) = \infty$, and so $\tilde{c}(p) \geq \tilde{c}(q)$. Now assume that $p_S(s) = \pi$. Then, since mutual information satisfies the data processing inequality (Cover and Thomas, 2006, Chapter 2), $MI(p) \geq MI(q)$ and so $\tilde{c}(p) \geq \tilde{c}(q)$. ■

Proof of Theorem 2.

The proof of (i) implies (ii) follows closely Theorem 1, Gilboa and Schmeidler (1989, Theorem 1) and Maccheroni et al. (2006, Proposition 19). We provide an outline of the basic steps below, omitting details that would be obvious.

Let $\succsim$ be a binary relation on $\mathcal{M}$ satisfying WO, VA, SI, MON, MC and UB, and define $\succsim^*$ on $\mathcal{K}_0$ by way of (1.18). Then by Lemma 1, $\succsim^*$ satisfies WO*, AA*, SI*, ΣMON*, MC* and UB*. Moreover, as in the proof of Theorem 1, each $K \in \mathcal{K}_0$ admits a constant-singleton equivalent $\mathbb{p}_K$ by MC*, and the preference relation $\succsim^*$ is represented by a convex functional I. It then follows as in Gilboa and Schmeidler (1989, Theorem 1, substituting sublinear for superlinear in the obvious way) and Lemma 29 in Maccheroni et al. (2006) that there exists an unbounded $u : \mathcal{Z} \to \mathbb{R}$, and a weak*-compact, convex set $\Pi \subset \Delta(\mathcal{S}\times\Sigma)$ such that $Eu[\mathbb{p}_K] = \max_{p\in\Pi} \int_{\mathcal{S}\times\Sigma}; (Eu \circ K) d\,p$. Moreover, Π is unique and u is unique up to positive linear transformations.

Also, since SI* implies WSI*, there exist a unique $\pi \in \Delta(\mathcal{S}\times\Sigma)$ and an information cost function $c^* : \Delta(\mathcal{S}\times\Sigma) \to [0,\infty]$ (relative to π) such that $Eu[\mathbb{p}_K] = \max_{p\in\Delta(\mathcal{S}\times\Sigma)} \int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\,p - c^*(p)$. Following exactly the steps in Proposition 19 of Maccheroni et al. (2006, substituting max for min in the obvious way), $c^*(p) \in \{0,\infty\}$ for all $p \in \Delta(\mathcal{S}\times\Sigma)$, and $Eu[\mathbb{p}_K] = \max_{p\in\{p'\in\Delta(\mathcal{S}\times\Sigma)|c^*(p')=0\}} \int_{\mathcal{S}\times\Sigma} (Eu \circ K)\, d\,p$.

By properties (a)-(d) of c^* in Theorem 1, we have: (i′) $c^*(p) = \infty$ for all $p \in \Delta(\mathcal{S}\times\Sigma)$ such that $p_{\mathcal{S}} \neq \pi$; (ii′) $c^*(p) = 0$ if $p \in \bar{\Delta}(\mathcal{S}\times\Sigma)$ and $p_{\mathcal{S}} = \pi$, (iii′) $c^*(q) = 0$ if $p \succsim q$ for some $p \in \Delta(\mathcal{S}\times\Sigma)$ such that $c^*(p) = 0$. Hence, the set $\{p' \in \Delta(\mathcal{S}\times\Sigma)|c^*(p') = 0\}$ is an information constraint set. Following the steps of Lemma 3 and the proof of (i) implies (ii) for Theorem 1, one concludes the sufficiency part of the proof.

To prove that (ii) implies (i), observe that $\tilde{V}$ is clearly a special case of V in Theorem 1, and hence $\succsim$ satisfies WO, VA, WSI, MON, MC and UB. It remains to show that $\succsim$ also satisfies SI. Let $\alpha \in (0,1)$, $M \in \mathcal{M}$, and $g \in \mathcal{F}$, then

$$\begin{aligned}
& \max_{p\in\Pi}\left(\int_\Sigma\left(\max_{f\in\alpha M+(1-\alpha)\{g\}}\int_{\mathcal{S}} Eu[f(s)]dp(s|\sigma)\right)dp_\Sigma(\sigma)\right) \\
= \; & \max_{p\in\Pi}\left(\int_\Sigma\left((\max_{h\in M}\int_{\mathcal{S}}\alpha Eu[h(s)]dp(s|\sigma)) + \int_{\mathcal{S}}(1-\alpha)Eu[g(s)]dp(s|\sigma)\right)dp_\Sigma(\sigma)\right) \\
= \; & \alpha\left[\max_{p\in\Pi}\left(\int_\Sigma\left(\max_{h\in M}\int_{\mathcal{S}} Eu[h(s)]dp(s|\sigma)\right)dp_\Sigma(\sigma)\right)\right] + (1-\alpha)\int_{\mathcal{S}} Eu[g(s)]d\pi(s)
\end{aligned} \tag{1.71}$$

Hence, $\tilde{V}(\alpha M + (1-\alpha)\{g\}) = \alpha\tilde{V}(M) + (1-\alpha)\tilde{V}(g)$, and so clearly $\succsim$ satisfies SI. ■

Proof of Proposition 2. Let $\pi \in \Delta(\mathcal{S}\times\Sigma)$, $\kappa \in [0,\infty]$ and define the sets $\Delta_\pi(\mathcal{S}\times\Sigma) =$

$\{p \in \Delta(\mathcal{S} \times \Sigma) | p_{\mathcal{S}} = \pi\}$ and $\tilde{\Pi} = \{p \in \Delta_\pi(\mathcal{S} \times \Sigma) | MI(p) \leq \kappa\}$. Clearly,

$$\tilde{V}^*(M) = \max_{\{p \in \tilde{\Pi}\}} \int_{\mathcal{S}} \left(\int_{\Sigma} \left(\max_{f \in M} \int_{\mathcal{S}} Eu[f(s)] \, d\, p(s|\sigma) \right) d\, p(\sigma|s) \right) d\, p_{\mathcal{S}}(s) \,, \tag{1.72}$$

so we only need to show that $\tilde{\Pi}$ is an information constraint set.

To see this, note that the mutual information functional $MI(.)$ is weak* lower semicontinuous in $p \in \Delta(\mathcal{S}\times\Sigma)$ and so, for any $\kappa \geq 0$, the set $\{p \in \Delta(\mathcal{S}\times\Sigma) | MI(p) \leq \kappa\}$ is weak* closed, and is therefore weak* compact. If, on the other hand, $\kappa = \infty$, then $\{p \in \Delta(\mathcal{S}\times\Sigma) | MI(p) \leq \kappa\} = \Delta(\mathcal{S}\times\Sigma)$, which is weak* compact. Also note that the set $\Delta_\pi(\mathcal{S}\times\Sigma)$ is weak* closed in $\Delta(\mathcal{S}\times\Sigma)$. Thus, $\tilde{\Pi}$ is a weak* closed subset of the set $\{p \in \Delta(\mathcal{S}\times\Sigma) | MI(p) \leq \kappa\}$, and hence $\tilde{\Pi}$ is weak* compact.

Next, we show that $\tilde{\Pi}$ is convex. Let $p, p' \in \tilde{\Pi}$ and $\alpha \in [0,1]$. Then the probability measure $q = \alpha p + (1-\alpha)p' \in \Delta_\pi(\mathcal{S}\times\Sigma)$. Moreover, by convexity of the mutual information functional,

$$\begin{aligned} MI(q) &\leq \alpha MI(p) + (1-\alpha) MI(p') \\ &\leq \alpha\kappa + (1-\alpha)\kappa = \kappa \,, \end{aligned}$$

showing that $q \in \tilde{\Pi}$.

Since $MI(p) = 0$ for any $p \in \bar{\Delta}(\mathcal{S}\times\Sigma)$, the set $\bar{\Delta}(\mathcal{S}\times\Sigma) \cap \Delta_\pi(\mathcal{S}\times\Sigma)$ is a subset of $\tilde{\Pi}$, showing that $\tilde{\Pi}$ satisfies property (i) of an information constraint set. Also note that by definition of $\tilde{\Pi}$, property (ii) is clearly satisfied.

Last, we verify property (iii). Let $p \in \tilde{\Pi}$ and $q \in \Delta(\mathcal{S}\times\Sigma)$ such that $p \succsim q$. By the proof of Proposition 1, $MI(p) \geq MI(q)$ and so $MI(q) \leq \kappa$, implying that $q \in \tilde{\Pi}$. This shows property (iii) and concludes the proof. ■

Proof of Theorem 3. (i) implies (ii): Let $\succsim$ be a variable attention preference relation that satisfies FA. By Theorem 1, $\succsim$ can be represented by a (π, u, c)-variable attention utility function $V : \mathcal{M} \to \mathbb{R}$. It is clear from the representation in (1.6) that for any $M, N \in \mathcal{M}$ and $\alpha \in (0, 1)$, $\max\{V(M), V(N)\} \geq V(\alpha M + (1 - \alpha)N)$. Thus, for any $M, N \in \mathcal{M}$ and $\alpha \in (0, 1)$,

$$M \succsim N \Rightarrow M \succsim \alpha M + (1 - \alpha)N \tag{1.73}$$

We now show that $\succsim$ must also satisfy SI. Let $M, N \in \mathcal{M}$, $f \in \mathcal{F}$ and $\alpha \in (0, 1)$.

We first show that $M \succsim N$ implies $\alpha M + (1 - \alpha)\{f\} \succsim \alpha N + (1 - \alpha)\{f\}$. Suppose that $M \succsim N$ and let $\mathbb{p}_M \in \Delta(Z)$ such that $M \sim \{\mathbb{p}_M\}$. By WO, $\{\mathbb{p}_M\} \succsim N$, and by FA and (1.73),

$$\begin{aligned} \alpha M + (1 - \alpha)\{\mathbb{p}_M\} &\sim \{\mathbb{p}_M\} \\ &\succsim \alpha N + (1 - \alpha)\{\mathbb{p}_M\}. \end{aligned} \tag{1.74}$$

Hence by (1.74) and WSI, $\alpha M + (1 - \alpha)\{f\} \succsim \alpha N + (1 - \alpha)\{f\}$.

To show the converse, suppose that $\alpha M + (1 - \alpha)\{f\} \succsim \alpha N + (1 - \alpha)\{f\}$ and let $\mathbb{p}_N \in \Delta(Z)$ such that $N \sim \{\mathbb{p}_N\}$. Then by FA and WSI,

$$\begin{aligned} \{\mathbb{p}_M\} &\sim \alpha M + (1 - \alpha)\{\mathbb{p}_M\} \\ &\succsim \alpha N + (1 - \alpha)\{\mathbb{p}_M\}. \end{aligned} \tag{1.75}$$

Thus, by (1.75), WSI and FA,

$$\begin{aligned} \alpha\{\mathbb{p}_M\} + (1 - \alpha)\{\mathbb{p}_N\} &\succsim \alpha N + (1 - \alpha)\{\mathbb{p}_N\} \\ &\sim \{\mathbb{p}_N\}. \end{aligned} \tag{1.76}$$

Hence, by (1.76) and WSI,

$$\{\mathbb{p}_M\} \succsim \alpha\{\mathbb{p}_N\} + (1 - \alpha)\{\mathbb{p}_M\}. \tag{1.77}$$

If $\alpha = 1/2$, by (1.76) and (1.77), $M \sim \{\mathbb{p}_M\} \succsim \{\mathbb{p}_N\} \sim N$, and so we are done. Otherwise, without loss of generality, assume that $\alpha > 1/2$ (the argument for

$\alpha < 1/2$ is analogous). Let $\beta = \frac{1-\alpha}{\alpha} \in (0,1)$. Using (1.76) and Step 1,

$$\begin{aligned}(1-\alpha)\{\mathbb{p}_M\} + \alpha\{\mathbb{p}_N\} &= \beta\left[\alpha\{\mathbb{p}_M\} + (1-\alpha)\{\mathbb{p}_N\}\right] + (1-\beta)\{\mathbb{p}_N\} \\ &\succsim \{\mathbb{p}_N\}. \end{aligned} \tag{1.78}$$

Hence, by (1.77) and (1.78), $\{\mathbb{p}_M\} \succsim \{\mathbb{p}_N\}$, so $M \succsim N$, and $\succsim$ satisfies SI.

Since $\succsim$ satisfies SI, it satisfies every premise of Theorem 2, there must exist a compact set $\Pi \subset \Delta(S \times \Sigma)$ such that the information cost function $c(.)$ satisfies:

$$c(p) = \begin{cases} 0 \text{ if } p \in \Pi \\ \infty \text{ otherwise} \end{cases}$$

Hence, for all $M \in \mathcal{M}$, we can write $V(M) = \max_{p\in\Pi} J(M,p)$, where the function $J(.,.)$ is defined in (1.46). Let $\Pi(M) = \arg\max_{p\in\Pi} J(M,p)$ for every $M \in \mathcal{M}$. Since Π is compact and J is continuous in p, $\Pi(M)$ is non-empty and closed. Moreover, if $M, N \in \mathcal{M}$ such that there exist some $\alpha > 0$ and $\beta \in \mathbb{R}$ satisfying $Eu \circ K_M = \alpha(Eu \circ K_N) + \beta$, then $\Pi(M) = \Pi(N)$.

We claim that $\bigcap_{M\in\mathcal{M}'} \Pi(M) \neq \emptyset$ for any finite $\mathcal{M}' \subset \mathcal{M}$. That is, the collection of sets $\{\Pi(M) \subset \Pi : M \in \mathcal{M}\}$ satisfies the finite intersection property. To see this, let $\mathcal{M}'$ be a finite subset of $\mathcal{M}$ and, in view of the preceding remark, assume w.l.o.g that $M \sim N$ for all $M, N \in \mathcal{M}'$. Enumerate the finite set $\mathcal{M}' = \{M_1, .., M_n\}$ for some $n \geq 1$. If $n = 1$, then claim holds trivially. Assume, therefore, that $n \geq 2$. Let $\alpha_i \in (0,1)$ for all $i \in \{1,2,...,n\}$ such that $(\sum_{i=1}^n \alpha_i) = 1$, and define $M = (\sum_{i=1}^n \alpha_i M_i)$. Since $M_i \sim M_j$ for all $i, j \in \{1,2,...,n\}$, by repeated application of FA, it follows that $M \sim M_i$ for all $i \in \{1,2,...,n\}$. Thus, for any $p \in \Pi(M)$,

$$\begin{aligned} V(M) &= J(M,p) = \sum_{i=1}^n \alpha_i J(M_i, p) \\ &\leq \sum_{i=1}^n \alpha_i V(M_i) = V(M) \end{aligned} \tag{1.79}$$

showing that $J(M_i, p) = V(M_i)$ for all $i \in \{1, 2, ..., n\}$ and so $p \in \bigcap_{M \in \mathcal{M}'} \Pi(M)$, establishing the claim.

Since Π is compact and the set $\{\Pi(M) \subset \Pi : M \in \mathcal{M}\}$ is a collection of closed sets in Π which satisfies the finite intersection property, it follows that $\bigcap_{M \in \mathcal{M}} \Pi(M) \neq \emptyset$ (see, e.g., Munkres, 1975, Theorem 5.9, p. 170). That is, there exists some $p \in \Pi$ such that $V(M) = J(M, p)$ for all $M \in \mathcal{M}$, showing that (ii) holds.

Moreover, if $q \in \Pi$ is another joint distribution such that $J(M, q) = V(M)$ for all $M \in \mathcal{M}$, then p and q must be "equally informative" Gollier (2004, Lemma 9 and Proposition 87, pp. 362–366). Thus, by Blackwell (1953)'s equivalence result, $p \simeq q$ (see Kihlstrom, 1984, Theorems 1–6, pp. 18–23).

(ii) implies (i): Let $\succsim$ be a binary relation on $\mathcal{M}$ represented by the function $\check{V} : \mathcal{M} \to \mathbb{R}$ given in (1.13). It is readily seen that $\check{V}$ is a (π, u, c)-variable attention utility function, and so $\succsim$ is a variable attention preference relation by Theorem 1. Moreover, since $\check{V}(M)$ is a linear functional on $\mathcal{M}$, $\succsim$ satisfies FA, and thus (i) holds. ■

Proof of Proposition 3. Let $\succsim_1, \succsim_2$ be preferences with variable attention on $\mathcal{M}$ represented by (u_1, π_1, c_1) and (u_2, π_2, c_2), respectively.

Suppose that $\succsim_2$ is more attentive than $\succsim_1$. By (1.14), for all $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$, $\{\mathbb{p}\} \sim_1 \{\mathbb{q}\}$ implies $\{\mathbb{p}\} \sim_2 \{\mathbb{q}\}$. We claim that we must also have $\{\mathbb{p}\} \succ_1 \{\mathbb{q}\}$ implies $\{\mathbb{p}\} \succ_2 \{\mathbb{q}\}$ for all $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$, and hence that $u_1 \approx u_2$. Assume for contradiction that there exist some $\mathbb{p}, \mathbb{q} \in \Delta(\mathcal{Z})$ such that $\{\mathbb{p}\} \succ_1 \{\mathbb{q}\}$ and $\{\mathbb{p}\} \sim_2 \{\mathbb{q}\}$. Since u_2 is non-constant, there exists some $\mathbb{r} \in \Delta(\mathcal{Z})$ such that either $\{\mathbb{r}\} \succ_2 \{\mathbb{p}\}$ or $\{\mathbb{q}\} \succ_2 \{\mathbb{r}\}$. Assume w.l.o.g that the former is true. By (1.14), it follows that $\{\mathbb{r}\} \succ_1 \{\mathbb{p}\}$ and so by MC, there exists some $\alpha \in (0,1)$ such that $\alpha\{\mathbb{r}\} + (1-\alpha)\{\mathbb{q}\} \sim_1 \{\mathbb{p}\}$. Then again by (1.14), it must follow that $\alpha\{\mathbb{r}\} + (1-\alpha)\{\mathbb{q}\} \sim_2 \{\mathbb{p}\} \sim_2 \{\mathbb{q}\}$. Since Eu_2 is affine on $\Delta(\mathcal{Z})$, it follows that

$$\begin{aligned} \alpha Eu_2[\mathbb{r}] + (1-\alpha)Eu_2[\mathbb{q}] &= Eu_2[\alpha\mathbb{r} + (1-\alpha)\mathbb{q}] \\ &= Eu_2[\mathbb{q}], \end{aligned} \tag{1.80}$$

showing that $Eu_2[\mathbb{r}] = Eu_2[\mathbb{q}]$ and so $\{\mathbb{r}\} \sim_2 \{\mathbb{q}\}$, a contradiction. This proves the claim.

We now show that $\pi_1 = \pi_2$. Assume for contradiction that this is not true. Then there must exists some $s \in S$ such that $\pi_1(s) \neq \pi_2(s)$. Also, since u_1 is non-constant, there exist some $\mathbb{q}, \mathbb{r} \in \Delta(\mathcal{Z})$ such that $\{\mathbb{q}\} \succ_1 \{\mathbb{r}\}$. Let $f \in \mathcal{F}$ such that $f(s) = \mathbb{q}$ and $f(s') = \mathbb{r}$ for all $s' \in S\backslash\{s\}$ and let $\mathbb{p}_f \in \Delta(\mathcal{Z})$ such that $\{\mathbb{p}_f\} \sim_1 \{f\}$. Then,

$$\pi_1(s)Eu_1[\mathbb{q}] + (1-\pi_1(s))Eu_1[\mathbb{r}] = Eu_1\left[\mathbb{p}_f\right],$$

and, since $u_1 \approx u_2$,

$$\pi_1(s)Eu_2[\mathbb{q}] + (1-\pi_1(s))\, Eu_2[\mathbb{r}] = Eu_2\left[\mathbb{p}_f\right]. \tag{1.81}$$

Also, by (1.14), we must have $\{\mathbb{p}_f\} \sim_2 \{f\}$ implying that

$$\pi_2(s)Eu_2[\mathbb{q}] + (1-\pi_2(s))\, Eu_2[\mathbb{r}] = Eu_2\left[\mathbb{p}_f\right]. \tag{1.82}$$

Combining (1.81) and (1.82), $Eu_2[\mathbb{q}] = Eu_2[\mathbb{r}]$ and so $Eu_1[\mathbb{q}] = Eu_1[\mathbb{r}]$, which implies that $\{\mathbb{q}\} \sim_1 \{\mathbb{r}\}$, a contradiction.

By the preceding arguments, it follows that $\pi_1 = \pi_2 = \pi$ and $u_1 \approx u_2$, so it is w.l.o.g to set $u_1 = u_2 = u$. Denote the niveloids corresponding to the representations of $\succsim_1$ and $\succsim_2$ on $\mathcal{K}_0$ in Lemma 3 by I_1 and I_2, respectively. By the supposition, if $M \sim_1 \{f\}$ for some $M, \{f\} \in \mathcal{M}$, then $M \succsim_2 \{f\}$, and therefore

$$I_1(Eu \circ K_M) = \int_{\mathcal{S}} Eu[f(s)]\, d\pi(s) \leq I_2(Eu \circ K_M)\,, \tag{1.83}$$

and so $I_1 \leq I_2$. Hence,

$$\begin{aligned} c_1(p) &= \sup_{K \in \mathcal{K}_0} \Big(\int (Eu \circ K)\, dp - I_1(Eu \circ K) \Big) \\ &\geq \sup_{K \in \mathcal{K}_0} \Big(\int (Eu \circ K)\, dp - I_2(Eu \circ K) \Big) = c_2(p)\,, \end{aligned} \tag{1.84}$$

for all $p \in \Delta(\mathcal{S} \times \Sigma)$, proving that (1.) implies (2.).

To prove the converse, suppose that $\pi_1 = \pi_2 = \pi$, $u_1 = u_2 = u$, and $c_1 \geq c_2$. Let $M, \{f\} \in \mathcal{M}$ such that $M \succsim_1 \{f\}$. Then by Theorem 1,

$$\max_{p \in \Delta(\mathcal{S} \times \Sigma)} \left(\int (Eu \circ K_M)\, dp - c_1(p) \right) \geq \int_{\mathcal{S}} Eu[f(s)]\, d\pi(s)\,. \tag{1.85}$$

But since $c_2 \leq c_1$, we have

$$\max_{p \in \Delta(\mathcal{S} \times \Sigma)} \left(\int (Eu \circ K_M)\, dp - c_2(p) \right) \geq \int_{\mathcal{S}} Eu[f(s)]\, d\pi(s)\,, \tag{1.86}$$

and so $M \succsim_2 \{f\}$ showing that (2.) implies (1.). ■

Proof of Corollary 1. By Proposition 3, $\succsim_2$ is more attentive than $\succsim_1$ iff $\pi_1 = \pi_2 = \pi$, $u_1 \approx u_2$, and $c_1 \geq c_2$ when $u_1 = u_2$.

(A) Let $\succsim_1, \succsim_2$ be preferences with constrained attention on $\mathcal{M}$ represented by (u_1, π_1, Π_1) and (u_2, π_2, Π_2), respectively. By the proof of Theorem 2, $\Pi_i = \{p \in \Delta(\mathcal{S} \times \Sigma) | c_i(p) = 0\}$ where $c_i : \Delta(\mathcal{S} \times \Sigma) \to \{0, \infty\}$, for constrained attention preferences $\succsim_i$, where $i = 1, 2$. Hence, setting $\pi_1 = \pi_2 = \pi$ and $u_1 = u_2 = u$, we have $\succsim_2$ is more attentive than $\succsim_1$ iff $c_1(p) = 0$ implies $c_2(p) = 0$, equivalently $\Pi_1 \subset \Pi_2$.

(B) Let $\succsim_1$ and $\succsim_2$ be preference relations with fixed attention on $\mathcal{M}$, represented by (π_1, u_1, p_1) and (π_2, u_2, p_2), respectively. By the proof of Theorem 3,

$\Lambda(p_i) = \{p \in \Delta(\mathcal{S} \times \Sigma) | c_i(p) = 0\}$ where $c_i : \Delta(\mathcal{S} \times \Sigma) \to \{0, \infty\}$, for fixed attention preferences $\succsim_i$, where $i = 1, 2$. Hence, setting $\pi_1 = \pi_2 = \pi$ and $u_1 = u_2 = u$, we have $\succsim_2$ is more attentive than $\succsim_1$ iff $c_1(p) = 0$ implies $c_2(p) = 0$, equivalently $\Lambda(p_1) \subset \Lambda(p_2)$ or $p_2 \succsim p_1$.

■

Proof of Proposition 4. In view of the preceding results, it is readily seen that

$$(vi) \Leftrightarrow (v) \Leftrightarrow (iv) \Leftrightarrow (iii) \Rightarrow (ii) \Rightarrow (i)\,.$$

We complete the proof by showing that *(i)* $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii)*.

First, suppose that (i) holds. Let $\succsim$ be preference relation with variable attention which satisfies SMON. For any menu $M \in \mathcal{M}$, observe that by (1.67) for every $f \in M$ and every $s \in \mathcal{S}$, $\{O_M(s)\} \succsim \{f(s)\}$ and thus $\{O_M\} \succsim M$ by SMON. Similary, by (1.67) for every $s \in S$ there exists some $f \in M$ such that $\{f(s)\} \sim \{O_M(s)\}$ and so $M \succsim \{O_M\}$ by SMON. Hence, SMON implies that $\{O_M\} \sim M$, and so $V(M) = V(\{O_M\})$ by Theorem 1. Moreover, since the information cost function $c(.)$ is grounded,

$$\begin{aligned} V(\{O_M\}) &= \int_S Eu[O_M(s)] d\pi(s) \\ &= \int_S \left(\max_{f \in M} Eu[f(s)] \right) d\pi(s) \end{aligned} \tag{1.87}$$

where the second equality follows from the definition of an oracle act in (1.67). Thus, (ii) holds.

Now assume that (ii) holds. By Theorem 1, for any $p \in \Delta(\mathcal{S} \times \Sigma)$

$$c(p) = \begin{cases} 0 \text{ if } p \in \bar{\Delta}_\pi(S \times \Sigma) \\ \infty \text{ if } p \notin \Delta_\pi(S \times \Sigma) \\ c \in [0, \infty] \text{ otherwise} \end{cases}$$

To show that c must be equal to 0, let $p \in \Delta_\pi(\mathcal{S} \times \Sigma)$. Take any $K' \in \mathcal{K}_0$. By (ii) it follows that $m(K') \sim \{O_{m(K')}\}$, and so $K' \sim^* K_{O_{m(K')}}$. Thus by Lemma 2,

$$\begin{aligned} I(Eu \circ K') &= I(Eu \circ K_{O_{m(K')}}) \\ &= \int_{\mathcal{S}} Eu\left[O_{m(K')}(s)\right] d\pi(s) \\ &= \int_{\mathcal{S}\times\Sigma} \left(Eu \circ K_{O_{m(K')}}\right) dp\,, \end{aligned} \tag{1.88}$$

where the last equality follows from $p_{\mathcal{S}} = \pi$. By definition of an oracle act, we see that $(Eu \circ K') \leq \left(Eu \circ K_{O_{m(K')}}\right)$, and so

$$\int_{\mathcal{S}\times\Sigma} (Eu \circ K')\, dp - \int_{\mathcal{S}\times\Sigma} \left(Eu \circ K_{O_{m(K')}}\right) dp \leq 0\,. \tag{1.89}$$

Thus, it follows from (1.45), (1.88) and (1.89) that $c(p) = 0$. This shows that (iii) holds and concludes the proof. ■

Proof of Proposition 5. In view of the preceding results, it is straightforward to see that

$$(vi) \Leftrightarrow (v) \Leftrightarrow (iv) \Leftrightarrow (iii) \Rightarrow (ii) \Rightarrow (i)\,.$$

We complete the proof by showing that *(i)* ⇒ *(ii)* ⇒ *(iii)*.

First, suppose that (i) holds; that is, let $\succsim$ be variable attention preference relation which satisfies the axiom NA. Consider any menu $M \in \mathcal{M}$. Since M is finite, there exists some $\succsim$-maximal element $g \in M$ such that $\{g\} \succsim \{f\}$ for all $f \in M$. By repeated application of NA, it follows that $\{g\} \sim M$, and so $V(M) = V(\{g\})$ by Theorem 1. Also, since the information cost function $c(.)$ is grounded, for any singleton menu $\{f\} \subset M$, $V(\{f\}) = \int_S Eu[f(s)]d\pi(s)$ by Theorem (1). Hence, as $\{g\} \succsim \{f\}$ for all $f \in M$, infer that $V(\{g\}) = \max_{f\in M} \int_S Eu[f(s)]d\pi(s)$, showing that (ii) holds.

Now suppose that (ii) holds. By Theorem 1, for any $p \in \Delta(\mathcal{S} \times \Sigma)$

$$c(p) = \begin{cases} 0 \text{ if } p \in \bar{\Delta}_\pi(S \times \Sigma) \\ \infty \text{ if } p \notin \Delta_\pi(S \times \Sigma) \\ c \in [0, \infty] \text{ otherwise} \end{cases}$$

In order to show that c must be equal to ∞, let $p \in \Delta_\pi(\mathcal{S}\times\Sigma)\backslash\bar{\Delta}_\pi(\mathcal{S}\times\Sigma)$. Then there must exist some signal $\sigma^* \in \Sigma$ such that the corresponding posterior distribution $p(.|\sigma^*) \neq \pi$. By definition of a conditional distribution, $p_\Sigma(\sigma^*) > 0$. Now let $f, g \in \mathcal{F}$ such that $\{f\} \succsim \{g\}$ and

$$\int_{\mathcal{S}} Eu[g(s)]dp(s|\sigma^*) > \int_{\mathcal{S}} Eu[f(s)]dp(s|\sigma^*)\,. \tag{1.90}$$

Let $K \in \mathcal{K}_0$ be a superact such that for all $(s, \sigma) \in \mathcal{S} \times \Sigma$,

$$K(s, \sigma) = \begin{cases} g(s) \text{ if } \sigma = \sigma^* \\ f(s) \text{ otherwise} \end{cases} \tag{1.91}$$

Thus, $m(K) = \{f, g\}$ and, since (ii) holds, $m(K) \sim \{f\}$. Therefore $K \sim^* K_f$, and so by Lemma 2,

$$\begin{aligned} I(Eu \circ K) &= I\left(Eu \circ K_f\right) \\ &= \int_{\mathcal{S}} Eu[f(s)]d\pi(s) \\ &= \int_{\mathcal{S}\times\Sigma} \left(Eu \circ K_f\right) dp\,, \end{aligned} \tag{1.92}$$

where the last equality follows from $p_S = \pi$. Thus, by (1.45), (1.92) and (1.90),

$$\begin{aligned} c(p) &\geq \int_{\mathcal{S}\times\Sigma} (Eu \circ K)dp - \int_{\mathcal{S}\times\Sigma} \left(Eu \circ K_f\right) dp \\ &= p_\Sigma(\sigma^*)\left[\int_S (Eu[g(s)] - Eu[f(s)])dp(s|\sigma^*)\right] > 0\,. \end{aligned} \tag{1.93}$$

Since u is unbounded, the right hand side of above inequality can be arbitrarily large. Hence, infer that $c(p) = \infty$, establishing that (iii) holds. ■

CHAPTER 2

ON REPRESENTATION OF MONOTONE PREFERENCES

2.1 Introduction

In this paper, we consider the problem of representability of *monotone* preference orders on a *sequence space*.[1] Monotone preferences are especially compelling in the theory of social evaluation of intertemporal utility streams since if no one is worse off, then the society as a whole should not be worse off (see Diamond (1965, p.172)). However, monotone preferences (expressing that "more is better") have also been used in the theory of individual preferences on commodity bundles, at least since the study of Wold (1943).[2]

The general characterization of representability of a preference ordering is the *order dense* property,[3]and this, of course, applies to our setting.[4] However, as is well known, the order dense property can be difficult to apply to concrete examples to decide on the representability (or non-representability) of a preference ordering. Thus, our objective is to present a sufficient condition for representability which provides a *partial characterization* but which is relatively easy to check in concrete examples.

We focus on a novel concept of *scalar continuity* of preferences, which may be

[1]We consider both finite and infinite sequence spaces as the theory developed here holds regardless of the dimension of the space.

[2]For a comprehensive discussion of Wold's result on the existence of a continuous utility function representing a preference order, see Beardon and Mehta (1994).

[3]Order dense property requires the existence of a countable subset which is dense in its given superset with respect to the underlying order topology. For a formal definition of this notion, see Remark 1, p.8.

[4]For expositions of this characterization result, see Fishburn (1970), Kreps (1988) and Bridges and Mehta (1995).

described as follows. Given any utility stream x, consider the set of *constant* utility streams which are at least as good as x, and the set of *constant* utility streams which are at most as good as x. These sets can be identified with sets of scalars, since constant utility streams are scalar multiples of the constant utility stream with constant utility equal to one. The preference order is scalar continuous if these sets of scalars are closed subsets of the real line. Thus scalar continuity requires continuity of preferences on the *diagonal* of the space of utility streams, and is therefore easy to verify. One of our main results (Proposition 7) is that a monotone, scalar continuous preference order can always be represented by a real valued function.

Here is a brief outline of the contents of the paper. Section 2.2 introduces the notation and definitions. In section 2.3, we associate with any monotone preference order a *pseudo utility function* μ which provides a *weak representation* of the order.[5] No continuity condition is imposed to obtain this pseudo utility function.

In section 2.4, we present our main representation results. In Proposition 7, we show that if the order satisfies a *scalar continuity* condition, then it is representable by the function μ.[6] Generalizing this result, we establish in Theorem 4 that when the set of equivalence classes (indifference curves) which have points of scalar discontinuity, is countable, then there exists a representation for the order. We indicate how *countable scalar discontinuity* condition can be used to verify the order dense property. However, this condition is not equivalent to the order dense property; an example is given to show that countable scalar discontinuity is *not* necessary for representability of a monotone preference order.

[5]This terminology follows Peleg (1970).

[6]This terminology follows Weibull (1985), who uses a similar concept of continuity.

In section 2.5, we consider a number of representation results which have been developed in the literature. We show that these results can be derived easily by applying Proposition 7 and Theorem 4.

In section 2.6, we provide two examples to illustrate the relationship between our concept of scalar continuity, and the concepts of sup-norm continuity and restricted continuity used in the literature.

2.2 Notation and Definitions

Let $\mathbb{N}$ denote, as usual, the set of natural numbers $\{1, 2, 3, ...\}$, and let $\mathbb{R}$ denote the set of real numbers. Let Y denote the closed interval $[0, 1]$, and let X denote the set Y^M where $M \in \mathbb{N} \cup \{\infty\}$.[7] Thus, $x \in X$ if and only if $x_n \in [0, 1]$ for all $n \in \mathbb{N}$ such that $n \leq M$ for some $M \in \mathbb{N} \cup \{\infty\}$. One can interpret x_n as the utility level of generation n, and x as an infinite stream of these utility levels or x_n as the consumption level of good n, and x as a finite bundle of these consumption levels.

The constant sequence of zeros in X will be denoted by 0, and the constant sequence of ones in X will be denoted by e. We denote the set of all constant sequences in X, $\{\lambda e \in X : \lambda \in [0, 1]\}$, by C and we call it the *diagonal* in X.

For $y, z \in \mathbb{R}^M$ with $M \in \mathbb{N} \cup \{\infty\}$, we write $y \geq z$ if $y_i \geq z_i$ for all $i \in \mathbb{N}$ such that $i \leq M$; $y > z$ if $y \geq z$, and $y \neq z$, and $y \gg z$ if $y_i > z_i$ for all $i \in \mathbb{N}$ such that $i \leq M$.

A *preference ordering* is a binary relation, $\succsim$ on X , which is complete and

[7]We use this notation to accommodate both the finite dimensional and infinite dimensional sequence spaces.

transitive. We associate with $\succsim$ its asymmetric and symmetric components by $\succ$ and $\sim$ respectively.

A preference ordering $\succsim$ on X is called *monotone* (M) if the following condition holds:

(M) If $x, y \in X$ and $x \geq y$, then $x \succsim y$.

We say that a preference ordering $\succsim$ on X is *strongly monotone* (SM) if it satisfies the following efficiency condition:

(SM) If $x, y \in X$, with $x > y$, then $x \succ y$.

When x and y are utility streams, then (SM) is the standard Pareto principle. Clearly the former efficiency condition is implied by the latter one. Strong monotonicity also implies the following condition called *weak Pareto*:

(WP) If $x, y \in X$ and $x \gg y$, then $x \succ y$.

A preference ordering $\succsim$ on X is *representable* if there is a function, $u : X \to \mathbb{R}$, such that for all $x, y \in X$,

$$x \succsim y \text{ if and only if } u(x) \geq u(y) \tag{R}$$

Given a preference ordering $\succsim$ on X, for each $x \in X$ the *Lower and Upper Contour Sets* are defined as $LC(x) = \{y \in X : x \succsim y\}$ and $UC(x) = \{y \in X : y \succsim x\}$ respectively.

Given a topology $\mathcal{T}$ for X, we say that $\succsim$ is $\mathcal{T}-$*continuous* on X if for each $x \in X$, the lower and upper contour sets ($LC(x)$ and $UC(x)$) of x are *closed* subsets of X in the topology $\mathcal{T}$.

2.3 Weak Representation of a Monotone Preference Order

We will associate with a monotone preference order $\succsim$ on X a *pseudo utility* function $\mu : X \to \mathbb{R}$ which provides a *weak representation* of it; that is,

$$\text{If } x, y \in X \text{ and } x \succsim y, \text{ then } \mu(x) \geq \mu(y) \tag{WR}$$

Condition WR implies that if $x, y \in X$ and $x \sim y$, then $\mu(x) = \mu(y)$. Also, if $x, y \in X$ and $\mu(x) > \mu(y)$, then $x \succ y$. However, it allows for the possibility that $x, y \in X$ satisfy $x \succ y$, but $\mu(x) = \mu(y)$. It is in this respect that the representation is weak.

For each $x \in X$ define the following subsets of the $[0, 1]$ interval:

$$A(x) = \{\lambda \in [0, 1] : \lambda e \succsim x\}; \; B(x) = \{\lambda \in [0, 1] : x \succsim \lambda e\} \tag{2.1}$$

Note that while upper and lower contour sets are subsets of X, the sets $A(x)$ and $B(x)$ are subsets of the real line.

Proposition 6 *Let $\succsim$ be a monotone preference ordering on X. Then, $\succsim$ has a weak representation.*

Proof. We obtain a weak representation as follows. For each $x \in X$, define $A(x)$ as in (2.1) and let $\alpha(x)$ be the infimum of the set $A(x)$. That is:

$$\alpha(x) = \inf_{\lambda \in A(x)} \lambda \text{ for each } x \in X \tag{2.2}$$

Note that since $\succsim$ is monotone, $A(x)$ is non-empty and so $\alpha(x)$ is well-defined. Clearly, $\alpha(x) \in [0,1]$ for each $x \in X$. We claim that $\alpha(x)$ satisfies condition WR. Let $x, y \in X$ satisfy $x \succsim y$. By (2.1), we have $A(x) \subset A(y)$ and thus $\alpha(x) \geq \alpha(y)$ by the monotonicity of an infimum. ■

Note that α is not the only possible weak representation of $\succsim$ on X. For instance, one can define $B(x)$ as in (2.1) and let $\beta(x)$ be the supremum of the set $B(x)$ for each $x \in X$. That is:

$$\beta(x) = \sup_{\lambda \in B(x)} \lambda \text{ for each } x \in X \tag{2.3}$$

Since $\succsim$ is monotone, $B(x)$ is non-empty and thus $\beta(x)$ is well-defined. Clearly, $\beta(x) \in [0,1]$ for each $x \in X$. Moreover, if $x, y \in X$ with $x \succsim y$, then $B(x) \supset B(y)$ by (2.1) and so $\beta(x) \geq \beta(y)$ by the monotonicity of a supremum. Therefore, β satisfies condition WR and is a weak representation of $\succsim$ on X.

In general, these two functions, α and β, need not be equal. However, $\alpha(x)$ can be at most $\beta(x)$ for each $x \in X$. For if $\alpha(x) > \beta(x)$ for some $x \in X$, then we can pick some $\theta \in (\beta(x), \alpha(x))$. Since $\succsim$ is complete and thus $A(x) \cup B(x) = [0,1]$ by (2.1), $\theta \in (0,1)$ must belong to $A(x)$ or $B(x)$. However, if $\theta \in A(x)$, we must have $\theta \geq \alpha(x)$ by (2.2), a contradiction. And, if $\theta \in B(x)$, we must have $\theta \leq \beta(x)$ by definition of $\beta(x)$, a contradiction. Thus we have:

$$\alpha(x) \leq \beta(x) \text{ for all } x \in X \tag{2.4}$$

In addition, by using α and β one can define many similar functions as well to serve as weak representations for the preference order $\succsim$. To see this, let $k \in (0,1)$ and define the function $\mu_k : X \to [0,1]$ as follows:

$$\mu_k(x) = k\alpha(x) + (1-k)\beta(x) \text{ for all } x \in X \tag{2.5}$$

Then if we let $x, y \in X$ with $x \succsim y$, we get $\mu_k(x) = k\alpha(x) + (1-k)\beta(x) \geq k\alpha(y) + (1-k)\beta(y) = \mu_k(y)$ exhibiting (WR).

We note here that if the preference order is monotone and satisfies Weak Pareto, then it can be shown that $\alpha(x) = \beta(x) = \mu_k(x)$ for all k in $(0, 1)$. However, monotone preference orders satisfying Weak Pareto need not be representable. The well-known example of the lexicographic preference order (see Debreu (1954)) satisfies Strong Monotonicity, and therefore is a monotone preference order satisfying Weak Pareto, but is not representable.[8]

2.4 Representation of a Monotone Preference Order

In this section, we use the weak representation result of Section 2.3 to provide a representation for monotone preference orders. For this purpose, we use a weak notion of continuity of preferences, called scalar continuity, to present our first representation result (Proposition 7). We then generalize this result to cover cases in which preference orders might exhibit a limited extent of scalar discontinuity (Theorem 4).

The following lemma is useful in obtaining our representation results.

Lemma 4 *Let $\succsim$ be a monotone preference ordering on X. Suppose x in X is a point*

[8]For a comprehensive study of complete preference orders which are not representable by a real valued function, see Beardon et al. (2002). As the characterization result in that paper indicates, these turn out to be of four types; an open question is which types can occur for complete preference orders which are monotone.

such that the sets $A(x)$ and $B(x)$ defined in (2.1) have a non-empty intersection. Then $\mu_k(x)e \sim x$ where μ_k is defined in (2.5).

Proof. Let x be in X be such that $A(x) \cap B(x)$ is non-empty. We have $\alpha(x) \leq \beta(x)$ by (2.4) and so we have two cases to consider; (i) $\alpha(x) < \beta(x)$, (ii) $\alpha(x) = \beta(x)$. In case (i), using (2.5) we get $\alpha(x) < \mu_k(x) < \beta(x)$. Then by (2.2) and (2.3), we have $\mu_k(x) \in A(x) \cap B(x)$ and so by (1), $x \sim \mu_k(x)e$. In case (ii), we have $\alpha(x) = \mu_k(x) = \beta(x)$ by (2.5). Thus given that $A(x) \cap B(x)$ is non-empty and using (2.2) and (2.3), for any $\lambda \in A(x) \cap B(x)$, we get $\alpha(x) \leq \lambda \leq \beta(x)$, and so $\lambda = \mu_k(x)$. This shows that $\mu_k(x) \in A(x) \cap B(x)$ and therefore by (2.1), we infer that $x \sim \mu_k(x)e$. ■

The following *weak* notion of continuity will be used in the next representation result.

Definition 5 *We say that $\succsim$ is* scalar continuous *on X if for each $x \in X$, the sets $A(x)$ and $B(x)$ defined in (2.1) are closed in the standard topology on $\mathbb{R}$.*

Sup-norm continuity (see Example 3 in Section 2.5) implies Restricted continuity (see Example 4 in Section 2.5), which in turn implies Scalar continuity. Further, there are monotone preference orders which satisfy Scalar continuity, but violate Restricted continuity (see Example 5 in the Appendix) and (therefore) Sup-norm continuity. Also, there are monotone preference orders which satisfy Restricted continuity but violate Sup-norm continuity (see Example 6 in the Appendix).

We can now state the following representation result for monotone preference orders.

Proposition 7 *If a monotone preference ordering $\succsim$ is scalar continuous on X, then $\succsim$ is representable, and μ_k, defined in (2.5), represents it.*

Proof. We know by (2.5) that μ_k is a weak representation of $\succsim$. Thus it remains to show that when $x, y \in X$ and $x \succ y$, we must have $\mu_k(x) > \mu_k(y)$.

Since $\succsim$ is scalar continuous, $A(x)$ and $B(x)$ are closed subsets in $[0, 1]$ for each $x \in X$. Moreover, since $\succsim$ in X is complete, then the union $A(x) \cup B(x)$ exhausts the interval $[0, 1]$ which is a connected set. Thus, $A(x) \cap B(x)$ must be non-empty for all $x \in X$. Thus, by Lemma 4 we have $\mu_k(x)e \sim x \succ y \sim \mu_k(y)e$ so that $\mu_k(x)e \succ \mu_k(y)e$ by transitivity. This implies $\mu_k(x) \neq \mu_k(y)$, and since by (WR), $\mu_k(x) \geq \mu_k(y)$, we must have $\mu_k(x) > \mu_k(y)$. ■

2.4.1 A Refinement of the Representation Result

For any preference order $\succsim$ on X, let:

$$D = \{x \in X : A(x) \cap B(x) = \emptyset\}$$

When $x \in D$, we refer to it as a *point of scalar discontinuity* of the preference order $\succsim$ on X. When D is empty, we say that the order $\succsim$ has no points of scalar discontinuity in X.

The ability to represent $\succsim$ depends crucially on (loosely speaking) "how many" points of scalar discontinuity there are. To make this notion precise, we make the preliminary remark that if $x \in X$ is a point of scalar discontinuity, and $y \sim x$, then y is also a point of scalar discontinuity. To see this, suppose $x \in D$. Since $y \sim x$, we have $A(x) = A(y)$ and $B(x) = B(y)$, and so $A(y) \cap B(y) = \emptyset$ implying that $y \in D$.

In view of this remark, proceed to form the following partition of D. For each $x \in D$, let $E(x) = \{z \in X : z \sim x\}$; clearly $E(x)$ is non-empty since $x \in E(x)$. For all $x, y \in X$, we have either $E(x)$ disjoint from $E(y)$, or $E(x) = E(y)$; further:

$$\bigcup_{x \in D} E(x) = D$$

Let $\mathfrak{I}$ be the collection $\{E(x)$ for some $x \in D\}$. Then, $\mathfrak{I}$ is a partition of D. In order to see "how much" scalar discontinuity the preference order exhibits, it is enough to look at "how many" equivalence classes there are in $\mathfrak{I}$. We can now introduce the following condition:

Countable Scalar Discontinuity Condition:

The collection $\mathfrak{I}$ has at most a countable number of equivalence classes.

We now show that this countable scalar discontinuity condition is *sufficient* for the representability of a monotone preference order.

Theorem 4 *If a monotone preference ordering $\succsim$ on X satisfies the countable scalar discontinuity condition, then $\succsim$ is representable.*

Proof. Let $\{F_1, F_2, ...\}$ be an enumeration of the set $\mathfrak{I}$. For each $F_n \in \mathfrak{I}$, let $r(F_n) = (1/2^n)$. Let $\mathbb{F}$ be the collection of all subsets of $\mathfrak{I}$ and define a function $\pi : \mathbb{F} \to \mathbb{R}$ as follows:

$$\pi(\mathfrak{I}') = \begin{cases} \sum_{F_n \in \mathfrak{I}'} r(F_n) & \text{if } \mathfrak{I}' \text{ is non-empty} \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

Note that since the sequence $\{r(F_n)\}$ is summable, π is well-defined, and indeed for any $\mathfrak{I}' \in \mathbb{F}$, $\pi(\mathfrak{I}') \in [0, 1]$.

For each $x \in X$, let $W(x) = \{y \in X : x \succ y\}$ and $P(x) = \{y \in X : y \succ x\}$. Note that for any $x \in X$, the sets $W(x)$ and $P(x)$ together separate $\mathfrak{I}$ into its subsets as for any $F_n \in \mathfrak{I}$, either $F_n \subset W(x)$ or $F_n \subset P(x)$ or $F_n \cap [W(x) \cup P(x)] = \emptyset$. For any $x \in X$, let $\mathfrak{I}_*(x) = \{F_n \in \mathfrak{I} : F_n \subset W(x)\}$ and $\mathfrak{I}^*(x) = \{F_n \in \mathfrak{I} : F_n \subset P(x)\}$, and define the function:

$$\rho(x) = \pi(\mathfrak{I}_*(x)) - \pi(\mathfrak{I}^*(x)) \tag{2.7}$$

Since π is bounded on $\mathbb{F}$, ρ is well-defined. Now consider the function $u : X \to \mathbb{R}$ defined as:

$$u(x) = \mu_k(x) + \rho(x) \tag{2.8}$$

where μ_k is defined in (2.5). We claim that u represents the ordering $\succsim$ on X.[9]

Let $x, y \in X$ such that $x \succsim y$. Then by the definition of the sets $\mathfrak{I}_*$ and $\mathfrak{I}^*$, we have $\mathfrak{I}_*(y) \subset \mathfrak{I}_*(x)$ and $\mathfrak{I}^*(x) \subset \mathfrak{I}^*(y)$. Thus by (2.6) and (2.7), we get $\rho(x) \geq \rho(y)$ and so by (2.5) and (2.8), $u(x) \geq u(y)$.

Now let $x, y \in X$ such that $x \succ y$. By (2.5) we have $\mu_k(x) \geq \mu_k(y)$. If $\mu_k(x) > \mu_k(y)$, then we have $u(x) > u(y)$ by (2.8) and the fact that $\rho(x) \geq \rho(y)$. If, however, $\mu_k(x) = \mu_k(y)$, then $x \in D$ or $y \in D$ must hold. Otherwise, by Lemma 4 we have $x \sim y$, a contradiction. In both cases, $x \in D$ or $y \in D$, one of the two set inclusions, $\mathfrak{I}_*(y) \subset \mathfrak{I}_*(x)$ and $\mathfrak{I}^*(x) \subset \mathfrak{I}^*(y)$, must be strict and so we must have $\rho(x) > \rho(y)$. Thus, by (2.8) $u(x) > u(y)$ which establishes the claim. ■

Remark 4 *Using the countable scalar discontinuity condition, one can directly check the* order dense property *for any monotone preference order; that is, one can find a*

[9]Our representation shows a link between measure and utility theory as the function π, defined in (2.6), is a simple measurable function. On a general approach using measure theory in constructing a representation, see Voorneveld and Weibull (2009).

countable subset Z which is order dense *in X.*[10] *To see this, let $Q = \{qe \in X :$ for some rational number $q\}$. Then $Q \subset X$ is countable. By countable scalar discontinuity, $\mathfrak{I}$ is countable and so by using the* Axiom of Denumerable Choice [11]*, one can pick an element from each $E \in \mathfrak{I}$; call this $g(E)$. Then the set $F = \{g(E) \in X : E \in \mathfrak{I}\}$ is also countable and so $Z = Q \cup F$ is a countable subset of X. We now verify that Z is order dense in X.*

Let $x, y \in X$, with $x \succ y$. There are two cases to consider; i) either x or $y \in D$ or ii) neither x nor $y \in D$. In case i), without loss of generality, let $x \in D$. Then $z = g(E(x)) \in F$ and $x \sim z$. Thus we have $x \succsim z \succsim y$. In case ii), we have $\mu_k(x) > \mu_k(y)$ by Lemma 4 and so there exists $z = qe \in Q$ such that $\mu_k(x) > q > \mu_k(y)$. By monotonicity of the order, we then have $\mu_k(x)e \succsim qe \succsim \mu_k(y)e$ and so $x \succsim z \succsim y$. This shows that Z is order dense in X.[12]

Remark 5 *A topic that has been discussed extensively in the social choice literature is the possible incompatibility of an efficiency concept like Strong Pareto with an equity concept like Anonymity when $X = Y^\infty$. Basu and Mitra (2003) showed that any preference order satisfying Strong Pareto and Anonymity cannot be represented by a real valued function. Further, although Svensson (1980) showed that preference orders satisfying Strong Pareto and Anonymity exist, the results of Zame (2007) and Lauwers (2010) imply that such preferences cannot be* constructed *and require the use of the Axiom of Choice or similar contrivance for demonstrating their existence.*

Our Theorem 4 implies that any preference order satisfying Strong Pareto and

[10] Z is order dense in X (in the sense of Debreu) if $x, y \in X$ and $x \succ y$ imply that there is some $z \in Z$, such that $x \succsim z \succsim y$. (See Bridges and Mehta (1995, p.11-12)).

[11] This axiom of set theory is a weak form of the Axiom of Choice. For an exposition of the Axiom of Choice, see (Munkres, 1975, p.59).

[12] Thus the result of Theorem 4 also follows by appealing to the order dense characterization result on representability if one grants the Axiom of Denumerable Choice.

Anonymity must have an uncountable number of equivalence classes, which have points of scalar discontinuity. Recall that if $x \in X$ is a point of scalar discontinuity, then there is no $\lambda \in [0, 1]$ such that λe is indifferent to x. [For, if there were such a λ, then this λ would belong to both $A(x)$ and $B(x)$, contradicting the fact that x is a point of scalar discontinuity]. Thus, there is an uncountable number of indifference curves, generated by such a preference order, which are disjoint from the diagonal of X. This provides further insight about the nature of efficient and equitable preference orders on infinite utility streams.

An Example

We now present an example in $X = Y^2$ to show that the countable scalar discontinuity condition is *not a necessary condition* for representability of a monotone preference order $\succsim$. This also shows that for monotone preference orders, countable scalar discontinuity is not equivalent to the order dense property. Thus, Theorem 4 is only a partial characterization of the representability of a monotone preference order.

Let us define $u(x_1, x_2)$ for all $(x_1, x_2) \in X = Y^2$ as follows:

$$u(x_1, x_2) = \begin{cases} x_1 & \text{for } x_1 \in [0, (1/2)) \\ (1/2) + x_2 & \text{for } x_1 = (1/2) \\ 1 + x_1 & \text{for } x_1 \in ((1/2), 1] \end{cases} \tag{2.9}$$

Then, define $\succsim$ on X as follows. For $x, y \in X$, $x \succsim y$ if and only if $u(x) \geq u(y)$. Then $\succsim$ is clearly a preference order, and it is monotone.

Let $U = \{x \in X : x_1 = (1/2) \text{ and } x_2 \in ((1/2), 1]\}$. Note that whenever $x \in U$, we have $A(x) = ((1/2), 1]$ and $B(x) = [0, (1/2)]$. Thus $D \supset U$. Moreover, for any $x, x' \in U$, we have $x \sim x'$ if and only if $x = x'$. Thus, $\mathfrak{I} \supset \{\{x\} : x \in U\}$ and so $\mathfrak{I}$ is uncountable. But, (2.9) is clearly a representation of $\succsim$ on X.

2.5 Applications of the Representation Result

We now consider four different representation results which have been derived in the literature. Our aim, here, is to show that each example below provides sufficient conditions to derive scalar continuity and/or countable scalar discontinuity conditions, and thereby to demonstrate that these well-known representation results follow from our results established in Section 2.4.

Example 1: Wold's Representation

The first theory on the existence of a continuous representation for a preference order was given in a fundamental paper of Wold (1943). Wold considers strongly monotone preference order, and establishes a representation by showing that every indifference class meets the diagonal given the condition below. We show how this result can be obtained by applying Proposition 7.

Let $X = Y^n$ for some $n \in \mathbb{N}$ and let $\succsim$ be a strongly monotone preference order on X satisfying the following continuity condition:

Wold: For any $x, y, z \in X$ such that $x \succ y \succ z$, there exist some $a, b \in (0, 1)$ such that $ax + (1 - a)z \succ y \succ bx + (1 - b)z$.

Using strongly monotone preferences, $x \sim 0$ if and only if $x = 0$, and $x \sim e$ if and only if $x = e$. Further $A(0) = [0, 1], B(0) = \{0\}$, and $A(e) = \{1\}, B(e) = [0, 1]$. Thus, for $x \sim 0$, $A(x)$ and $B(x)$ are closed sets; and for $x \sim e$, $A(x)$ and $B(x)$ are closed sets. Consider then any $x \in X$, such that $e \succ x \succ 0$. We show that $A(x)$ is closed as follows. Let $\{\lambda^s\}_{s=1}^{\infty}$ be a convergent sequence of elements in $A(x)$, converging to λ^0. We have to show that $\lambda^0 \in A(x)$. If this is not the case, then we have $\lambda^0 \in [0,1]$, and $e \succ x \succ \lambda^0 e$. So, $\lambda^0 < 1$, and by Wold's condition, there is $b \in (0,1)$, such that $x \succ [b\lambda^0 + (1-b)]e$. Since $\lambda^s \to \lambda^0$ as $s \to \infty$ and $\lambda^0 < [b\lambda^0 + (1-b)]$, we can find s' large enough for which $\lambda^{s'} < [b\lambda^0 + (1-b)]$. But, then, $x \succ \lambda^{s'} e$, a contradiction to the fact that $\lambda^s \in A(x)$ for all $s \in \mathbb{N}$. Thus, $\lambda^0 \in A(x)$, and $A(x)$ is closed. We can show that $B(x)$ is closed by following a similar line of proof. Thus, the preference order $\succsim$ satisfies scalar continuity and is representable, by using Proposition 7.

If the preference order is monotone but not strongly monotone, the sets $A(x)$ and $B(x)$ need not be closed for all $x \in X$. Consider the preference relation $\succsim$ on X for which $x \sim e$ if $x_n > 0$ for some $n \in \mathbb{N}$, and $e \succ 0$. This is easily seen to be a preference order, which is monotone. Further, Wold's condition is trivially satisfied since one cannot find three points x, y, z in X satisfying $x \succ y \succ z$. Now, if $x \in X$ with $x \neq 0$, then $A(x) = (0,1]$, so $A(x)$ is not closed. Thus, $\succsim$ does not satisfy scalar continuity, and Proposition 7 is not applicable.

However, if the preference order is monotone, and Wold's condition is satisfied, we can still show that the preference order is representable by using Theorem 4 as follows. By using Wold's condition, one can show the following fact (see Fishburn (1970, p.33)).

Fact: If $x, y, z \in X$ such that $x \succ y \succ z$, then there exists some $a \in [0,1]$ such that $y \sim ax + (1-a)z$.

We claim that there is no point of scalar discontinuity of $\succsim$ and thus by Theorem 4, $\succsim$ is representable. To see this, let $y \in X$. We have $e \geq y \geq 0$. If $y = 0$ or $y = e$, then $A(y) \cap B(y)$ is non-empty. If $y \neq 0$ and $y \neq e$, then $e > y > 0$. Then by the fact above, there is some a in $[0, 1]$ such that $y \sim ae + (1 - a)0 = ae$ showing that $A(y) \cap B(y)$ is non-empty for any $y \in X$.

Example 2: Weighted Utilitarian Representation

In the theory of social choice, one of the prominent judgment criteria on the welfare of the society is called (weighted) utilitarianism. This method seeks to maximize the society's collective welfare obtained by summing (weighted) individual utilities (see d'Aspremont and Gevers (2002) for a discussion of available characterization results). We show below how one can achieve the existence of a representation for a preference order satisfying a set of axioms used in this literature.

Let $X = Y^n$ for some $n \in \mathbb{N}$ and let $\succsim$ be a monotone preference order on X satisfying WP and the following two conditions:

Minimal Individual Symmetry (MIS): For all $i, j \in \{1, 2, ..., n\}$, there exist $x, y \in X$ such that $x_i > y_i$, $x_j < y_j$, $x_k = y_k$ for all $k \in \{1, 2, ..., n\} \backslash \{i, j\}$, and $x \sim y$.

Strong Invariance (SI): For all $x, y \in X$, $x \succsim y$ implies that for all $z \in \mathbb{R}^n$ and all $b \in \mathbb{R}_{++}$, we have $(bx + z) \succsim (by + z)$ whenever $(bx + z), (by + z) \in X$.

We claim that $\succsim$ is scalar continuous on X. Let $I = \{1, 2, 3, ..., n\}$. By using MIS, SI and WP, we can easily find a unique vector $(q_i)_{i \in I} \gg 0$ such that $e_i \sim q_i e$ for each $i \in I$. Thus by using SI, we can infer that for any $x \in X$, $x \sim \lambda(x)e$ where $\lambda(x) = \sum_{i \in I} q_i x_i$. Note that for all x in X we have $\lambda(x) \in [0, 1]$ since $\sum_{i \in I} q_i = 1$ and

$x_i \in [0, 1]$ for every $i \in I$, and thus $\lambda(x) \in A(x) \cap B(x)$.

Since the preference order satisfies WP, we can infer that $\alpha(x) = \beta(x)$ for all x in X. To see this, let $\alpha(x) < \beta(x)$. Define $\varepsilon = \beta(x) - \alpha(x)$, and find $\delta \in (0, (\varepsilon/2))$, $\mu \in A(x)$ and $\eta \in B(x)$ such that $\mu < \alpha(x)+\delta$ and $\eta > \beta(x)-\delta$. Then, $\eta > \beta(x)-(\varepsilon/2) = [\beta(x) - \alpha(x)] + \alpha(x) - (\varepsilon/2) = \alpha(x) + (\varepsilon/2) > \mu$ so that we have, using the fact that $\succsim$ is monotone $x \succsim \eta e > \mu e \succsim x$ contradicting the transitivity of $\succsim$.

Since $\lambda(x) \in A(x) \cap B(x)$, we have a non-empty intersection of $A(x)$ and $B(x)$ and thus we must have $\lambda(x) = \alpha(x) = \beta(x)$ for each $x \in X$. Thus $\alpha(x) \in A(x)$ and $\beta(x) \in B(x)$ and so $A(x) = [\lambda(x), 1]$ and $B(x) = [0, \lambda(x)]$ by (2.2) and (2.3). This shows that $A(x)$ and $B(x)$ are closed in Y and therefore the preference order is scalar continuous, and has a representation by using Proposition 7.

Mitra and Ozbek (2010) show that whenever a preference order satisfies MIS, Invariance (a weaker form of SI), WP, and has a representation, then it also has a weighted utilitarian representation. Thus Proposition 7, together with the conditions above, ensures a weighted utilitarian representation for the preference ordering.

Example 3: Diamond's Representation on Infinite Utility Streams

The framework for analysis of social preference orders on infinite utility streams was introduced by Koopmans (1960). Diamond (1965) established the existence of a representation for monotone preference orders, which satisfy weak Pareto and sup-norm continuity. We now show how his existence result can be derived from Proposition 7.[13]

[13]This approach coincides with Yaari's given in a footnote in Diamond (1965).

Let $X = Y^\infty$ and let $\succsim$ be a monotone preference order on X satisfying WP, and:

Sup-Norm Continuity: For each $x \in X$, the Lower and Upper Contour Sets, $LC(x) = \{y \in X : x \succsim y\}$ and $UC(x) = \{y \in X : y \succsim x\}$ respectively, are closed with respect to the sup-norm.

We claim that $\succsim$ satisfies scalar continuity. Let $x \in X$. Then the two sets $UC(x) \cap C$ and $LC(x) \cap C$ are closed in the sup-norm topology since C, the set of constant sequences in X, is also a closed set in the sup-norm topology. One can easily verify that the function $\pi : [0,1] \to C$, defined as $\pi(k) = ke$ for every $k \in [0,1]$, is continuous. Then we have $A(x) = \pi^{-1}(UC(x) \cap C)$ and $B(x) = \pi^{-1}(LC(x) \cap C)$ showing that $A(x)$ and $B(x)$ are closed sets in $[0,1]$. Thus $\succsim$ satisfies scalar continuity and so $\succsim$ is representable by Proposition 7.

Example 4: Asheim-Mitra-Tungodden Representation on Infinite Utility Streams

Preference orders on infinite utility streams, which can be represented, exhibit a conflict if the order is required to satisfy certain equity and efficiency axioms simultaneously. (See Diamond (1965), Basu and Mitra (2003), Hara et al. (2008), and others).

Seeking a way out of such impossibility results, Asheim et al. (2012) introduce weak versions of efficiency and equity, together with a weak continuity requirement to establish a class of sustainable recursive social welfare functions for monotone preference orders. In doing this, they first show the existence of representation for the preference orders by using a continuity condition called

"Restricted Continuity" which is weaker than the usual sup–norm continuity. We show here how one can derive the existence of a representation by appealing to Proposition 7.

Let $X = Y^\infty$ and let $\succsim$ be a monotone preference order on X satisfying the following continuity condition:

Restricted Continuity: For all $x, y \in X$, if x satisfies $x_t = z$ for all $t > 1$, and the sequence streams $\{x^n\}_{n\in\mathbb{N}}$ satisfies $\lim_{n\to\infty} \sup_t |x^n_t - x_t| = 0$ with, for each $n \in \mathbb{N}$, $x^n \succsim y$ (resp. $y \succsim x^n$), then $x \succsim y$ (resp. $y \succsim x$).

We claim that $\succsim$ satisfies scalar continuity. Let $x \in X$ and consider set $A(x)$. By definition of $\alpha(x)$, there exists a sequence $\{\lambda^s\}_{s\in\mathbb{N}}$ in $A(x)$ such that $\lambda^s \to \alpha(x)$ as $s \to \infty$. Define $x^s = \lambda^s e \in X$ for all $s \in \mathbb{N}$. Then by definition of set $A(x)$, we have $x^s \succsim x$ for all $s \in \mathbb{N}$. Moreover, as $s \to \infty$, $x^s \to \alpha(x)e$ in sup-norm metric. Thus by restricted continuity $\alpha(x)e \succsim x$ and so $\alpha(x) \in A(x)$ using definition of set $A(x)$. Since the order $\succsim$ is monotone, we must have $A(x) = [\alpha(x), 1]$ which is a closed set in $[0, 1]$. Following a similar argument for set $B(x)$, one can show that $B(x) = [0, \beta(x)]$ which is also a closed set in $[0, 1]$. Thus $\succsim$ satisfies scalar continuity and so by Proposition 7, $\succsim$ is a representable preference order.

2.6 Conclusion

In this paper, we have investigated the relation between scalar continuity and representability of monotone preference orders in a sequence space. Scalar continuity is shown to be sufficient for representability of a monotone preference order (Proposition 7). Generalizing this result, we have shown that a countable scalar discontinuity condition ensures representability of a monotone preference order (Theorem 4). Although these conditions are not necessary conditions for representability of a monotone preference order, they are very useful for applications. We have demonstrated this by indicating how some of the well-known representation results from the literature follow from our representation results established in Proposition 7 and Theorem 4. Moreover, we have related the countable scalar discontinuity condition to the well-known order dense property (Remark 4), which is both necessary and sufficient for representability.

2.7 Appendix

In this section, we show that (i) there are monotone preference orders which satisfy Scalar continuity, but violate Restricted continuity and (therefore) Sup-norm continuity (Example 5); and (ii) there are monotone preference orders which satisfy Restricted continuity but violate Sup-norm continuity (Example 6).

2.7.1 Scalar but not Restricted Continuous Monotone Preference Order

Example 5: We first construct the example without specifying any dimension for the sequence space, but we later indicate which case of it we are using, finite or infinite, when we are considering the relevant conditions for that case.

Let $(q_n)_{n \leq M}$ be a sequence for some $M \in \mathbb{N} \cup \{\infty\}$ satisfying:

$$q_n > 0 \text{ for all } n \leq M \text{ and } \sum_{n=1}^{M} q_n = 1 \tag{2.10}$$

Let $f : X \to \mathbb{R}$ be defined by:

$$f(x) = \begin{cases} \sum_{n=1}^{M} q_n x_n & \text{if } x_n > 0 \text{ for all } n \leq M \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

Note that the sum in the definition of f in (2.11) converges for any $x \in X$ and thus f is well-defined. Define $\succsim$ by:

$$\text{For all } x, y \in X, \;\; x \succsim y \;\text{ if and only if } f(x) \geq f(y) \tag{2.12}$$

Then, $\succsim$ is a preference ordering on X, and f is a real-valued representation of it. Since f is increasing, [that is, for any $x, y \in X$, $x \geq y$ implies $f(x) \geq f(y)$ and $x \gg y$ implies $f(x) > f(y)$], by (2.12) we have for any $x, y \in X$, if $x \geq y$, then $x \succsim y$ and if $x \gg y$, then $x \succ y$ and thus, $\succsim$ is monotone and it also satisfies WP.

Scalar Continuity: Let $x \in X$. Then by (2.11) and (2.12), $A(x) = [f(x), 1]$ and $B(x) = [0, f(x)]$ which are both closed sets in $[0, 1]$ showing that $\succsim$ satisfies scalar continuity.

Restricted Continuity: We now show that $\succsim$ does not satisfy restricted continuity on $X = Y^{\infty}$. Let $z = (1 - q_1)e$ and consider the sequence $\{x^n\}_{n=1}^{\infty}$ in X where $x^n = (\frac{1}{n}, 1, 1, 1, ...)$ for all $n \in \mathbb{N}$. Then for all $n \in \mathbb{N}$, we have $f(x^n) = (1 - (\frac{n-1}{n})q_1) \geq (1 - q_1) = f(z)$ implying by (2.12) that:

$$x^n \succsim z \text{ for all } n \in \mathbb{N} \tag{2.13}$$

Let $x = (0, 1, 1, 1, ...)$ in X and note that:

$$d(x^n, x) = \frac{1}{n} \to 0 \text{ as } n \to \infty \tag{2.14}$$

We have by (2.11), $f(x) = 0 < (1 - q_1) = f(z)$, and hence by (2.12):

$$z \succ x \tag{2.15}$$

But then (2.13), (2.14) and (2.15) imply that $\succsim$ on Y^∞ violates the restricted continuity condition.

Sup-Norm Continuity: Since continuity of an order in sup-norm topology implies restricted continuity, we infer from the result above that $\succsim$ on $X = Y^\infty$ does not satisfy sup-norm continuity.

We now observe that Example 1 also demonstrates that representation of the preference order in this example cannot be obtained by applying the representation results of Wold (1943), Diamond (1965), d'Aspremont and Gevers (2002), and Asheim et al. (2012).

Since in this example, restricted continuity and therefore sup-norm continuity, is violated, the representation results of Asheim et al. (2012) and Diamond (1965) are not applicable. We verify below that the example also violates Wold's continuity condition and Strong Invariance, so that the representation results of Wold (1943) and d'Aspremont and Gevers (2002) are also not applicable.

Wold: Let $\succsim$ be an order on $X = Y^n$ defined in (2.12) for some given $n \in \mathbb{N}$. Let $x = e \in X$, $y = (1 - q_1)e \in X$ and $z = (e - e_1) \in X$. Then, by (2.11), we have $f(x) = 1 > (1 - q_1) = f(y) > f(z) = 0$ and so by (2.12), $x \succ y \succ z$. Let $a \in (0, 1)$ and define $w(a) = ax + (1 - a)z$. We have $w(a) = (e - (1 - a)e_1)$. By (2.11), we get $f(w(a)) = (1 - (1 - a)q_1) > (1 - q_1) = f(y)$ and so by (2.12), $w(a) \succ y$. This shows that there is no $b \in (0, 1)$ such that $y \succ w(b)$. Thus, the order $\succsim$ does not satify the condition of Wold.

Strong Invariance: Let $\succsim$ be an order on $X = Y^n$ defined in (2.12) for some given $n \in \mathbb{N}$. Let $x = (e - e_1) \in X$, $y = ((1 - q_1)/2)e \in X$, $b = 1$ and $z = (1/2)e_1 \in \mathbb{R}^n$. Then we have $(x + z) = (e - (1/2)e_1) \in X$ and $(y + z) = ((1 - q_1)/2)e + (1/2)e_1 \in X$. Moreover, by (2.11) $f(y) = ((1-q_1)/2) > 0 = f(x)$ and so by (2.12), $y \succ x$. Similarly, by (2.11), $f(x + z) = 1 - (1/2)q_1 > ((1 - q_1)/2) + (1/2)q_1 = f(y + z)$ and so by (2.12), $(x + z) \succ (y + z)$ showing that SI is not satisfied.

2.7.2 Restricted but not Sup-norm Continuous Monotone Preference Order

Example 6: Let $X = Y^\infty$ and consider a sequence $(q_n)_{n \in \mathbb{N}}$ defined as in (2.10). Let $g : X \to \mathbb{R}$ be defined by:

$$g(x) = \begin{cases} \sum_{n=1}^{\infty} q_n x_n & \text{if } x_n > 0 \text{ for all } n > 1 \\ q_1 x_1 & \text{otherwise} \end{cases} \tag{2.16}$$

Note that the sum in the definition of g in (2.16) converges for any $x \in X$ and thus g is well-defined. Define $\succsim$ by:

$$\text{For all } x, y \in X, \ x \succsim y \ \text{ if and only if } g(x) \geq g(y) \tag{2.17}$$

Then, $\succsim$ is a preference ordering on X, and g is a real-valued representation of it. Since g is increasing, [that is, for any $x, y \in X$, $x \geq y$ implies $g(x) \geq g(y)$ and $x \gg y$ implies $g(x) > g(y)$], by (2.17) we have for any $x, y \in X$, if $x \geq y$, then $x \succsim y$ and if $x \gg y$, then $x \succ y$ and thus, $\succsim$ is monotone and it also satisfies WP.

Sup-Norm Continuity: We first show that $\succsim$ does not satisfy sup-norm continuity on $X = Y^{\infty}$. Let $z = (1 - q_2)e$ and consider the sequence $\{x^n\}_{n=1}^{\infty}$ in X where $x^n = (1, \frac{1}{n}, 1, 1, ...)$ for all $n \in \mathbb{N}$. Then for all $n \in \mathbb{N}$, we have $g(x^n) = (1 - q_2) + (1/n)q_2 \geq 1 - q_2 = g(z)$ implying by (2.17) that:

$$x^n \succsim z \text{ for all } n \in \mathbb{N} \tag{2.18}$$

Let $x = (1, 0, 1, 1, ...)$ in X and note that:

$$d(x^n, x) = \frac{1}{n} \rightarrow 0 \text{ as } n \rightarrow \infty \tag{2.19}$$

We have by (2.10) and (2.16), $g(x) = q_1 < 1 - q_2 = g(z)$, and hence by (2.17), we get:

$$z \succ x \tag{2.20}$$

But then (2.18), (2.19) and (2.20) imply that $\succsim$ on Y^{∞} violates the sup-norm continuity condition.

Restricted Continuity: We now show that the order $\succsim$ satisfies the restricted continuity condition. To see this, let $y \in X$ and $\{x^n\}_{n=1}^{\infty}$ be a sequence in X with $x^n \succsim y$ for all $n \in \mathbb{N}$ converging in sup-norm to some $x \in X$ such that for all $m > 1$, $x_m = a$ for some $a \in [0, 1]$ (A similar line of argument can be given for the case where $y \succsim x^n$ for all $n \in \mathbb{N}$). There are two cases to consider: either (i) there exists $N \in \mathbb{N}$ such that for every $n \geq N$, $x^n_m > 0$ for all $m > 1$ or (ii) for every $N \in \mathbb{N}$ there exists $n \geq N$ such that $x^n_{m(n)} = 0$ for some $m(n) \in \mathbb{N}$ with $m(n) > 1$.

In case (i), by (2.16) and (2.17) we have $g(x^n) = qx^n \geq g(y)$ for all $n \geq N$. Since qx is sup-norm continuous, we have $qx^n \rightarrow qx$ as $n \rightarrow \infty$ and so $qx \geq g(y)$. Note that since for all $m > 1$, $x_m = a$ for some $a \in [0, 1]$, we have $g(x) = qx$ and so we must have $g(x) \geq g(y)$ implying that $x \succsim y$.

In case (ii), we can find a subsequence $\{x^{n_k}\}_{k=1}^{\infty}$ such that $x^{n_k}_{m_k} = 0$ for some $m_k \in \mathbb{N}$ with $m_k > 1$. Thus for all $k \in \mathbb{N}$, by (2.16) and (2.17) we have $g(x^{n_k}) = q_1 x_1^{n_k} \geq g(y)$. Since we have $x^{n_k} \to x$ as $k \to \infty$ in sup-norm, we must have $x_1^{n_k} \to x_1$ as $k \to \infty$ and thus $q_1 x_1^{n_k} \to q_1 x_1$ as $k \to \infty$. Therefore, by (2.10), (2.16) and (2.17) we get $g(x) \geq q_1 x_1 \geq g(y)$ inferring that $x \succsim y$.

Scalar Continuity: Note that since the monotone order $\succsim$ on $X = Y^{\infty}$ satisfies restricted continuity, it must also satisfy scalar continuity following the discussion given in Example 4 of Section 2.5.

CHAPTER 3

RATIONAL INATTENTION AND CHOICE OF OPTIMAL INFORMATION

3.1 Introduction

In a series of pioneering works, Sims (1998, 2003) introduced a model of information acquisition according to which the decision maker (DM) might disregard valuable information to save from the costs. That is, the DM might rationally choose not to be attentive to all the available information due to her attention costs among many others. As a result, the DM makes decisions often with imperfect knowledge leading to many interesting behavioral implications where a perfectly rational model lacks to predict. Moreover, it is shown in different choice contexts that rational inattention is generally applicable unlike many models of bounded rationality.

Indeed, rational inattention models have been widely used in the literature to provide plausible answers to some existing questions and new insights on many others.[1] To keep the analysis tractable, however, most of the applied work assume specific distributions on the *state-action* set as a way of representing the available information. As it is argued in detail before, for instance by Sims Sims (2006), these assumptions are not only restrictive but also hard to defend.[2]

[1]Models with rationally inattentive agents have been applied to study *inter alia* microeconomic foundations for price stickiness, Sims (1998, 2003), investment inertia, Woodford (2009), coordination failures,Hellwig and Veldkamp (2009)), business-cycle dynamics,Mackowiak and Wiederholt (2010), discrete pricing dynamics, Matějka (2012), and under-diversification of portfolios,Van Nieuwerburgh and Veldkamp (2010). See Veldkamp (2011) for an exhaustive survey of the rational inattention literature.

[2]In this regard, most applications of rational inattention theory consider normally distributed state-signal structures. However, it should be noted that a normally distributed signal structure need not necessarily be optimal unless a linear-quadratic payoff function is used. See Sims (2006) and Matějka and McKay (2013).

As an attempt to overcome these shortcomings, this work considers an abstract choice setting and shows that optimal choices of information can be found without use of any restrictions on the type of available information. Despite its mathematical novelty, this result I believe is quite important from an economic point of view. This is due to the fact that models of rational inattention with a fully endogenous choice of information should and would lead to a richer set of behavioral implications and thereby explain more of the observed economic phenomena.

In specific, this work considers a set of acts available to the DM from which she eventually makes a choice. Each act gives a payoff to the DM depending on a finite number of states of the world. The DM, on the other hand, is uncertain about which the true state is but as a Bayesian, she can process available information to make better choices. To this regard, her information processing activities, called *information channels*, are modeled as a distribution over the set of posteriors consistent with Bayesian updating. Each information channel is costly and the more *informative* they become, the more costly they get. Therefore, the DM needs to make a tradeoff in choosing a channel defining the information processing problem of the DM.

Theorem 5 in section 3.2.2 then shows that when the cost of processing information is *linear* in channels, a form admitted by the cost functions applied in the literature, one can obtain an optimal channel with a *finite* support. Indeed, it is shown that these optimal posteriors have to be the extreme points of a hyperplane supporting the expected net-utility frontier at the prior (see Figure 3.2 in section 3.3.2). Therefore, solving for an optimal channel amounts to first finding a finite number of posteriors by appealing to standard optimization techniques and then forming a channel in a consistent manner with Bayesian updating.

Moreover, as an application of Theorem 5, a simple buyer-seller problem is considered in section 3.3. In this problem, which has many applications in real-life, there is a seller who needs to price a risky project to offer to a potential buyer who suffers from costly attention. In Proposition 9, the buyer's unique optimal channel is obtained with its support and shown to satisfy many intuitive properties. For instance, when the offer price is low enough relative to the cost of information processing, the DM does not process any information at all. That is, she uses only her prior in accepting the offer. In contrast, when the offer price is relatively high enough, she rejects the offer without again any use of information. In between these extreme prices, she processes some information and depending on the realization of a posterior she either accepts or rejects the offer.

Proposition 10 then gives the buyer's acceptance likelihood of the offer. It is shown that the buyer becomes more likely to accept the offer as the price of the offer gets lower. Thus, the seller faces a tradeoff between increasing the revenue he obtains from the offer if accepted and increasing the likelihood of the buyer accepting the offer. Finally, it is shown in Proposition 11 that, perhaps surprisingly, the seller's optimal pricing strategy should be *non-monotonic* in the degree of the buyer's cost of information processing.

The remainder of the paper is organized as follows. This section closes with a brief discussion of the related literature. Section 2 then formally describes the choice problem of a rationally inattentive DM and gives a characterization for the optimal information channels. In section 3, the buyer-seller problem is introduced and then analyzed by first deriving the buyer's unique optimal choice of information and the seller's optimal pricing strategies on certain domains. Section 4 then considers some future research questions in line of the buyer-seller problem and concludes. The proofs of all results are separately provided in an Appendix.

Related Literature

In applying models of rational inattention, it is common to focus on the stochastic choice of acts as a result of the DM's information processing activities. Therefore, it is common to model the DM's information choices as choosing joint distributions between the states and the acts. In this work, however, an alternative but an equivalent model of information choice is used following the axiomatic study of de Oliveira et al. (2013). This way of modeling, by understanding that there are optimal acts for each posterior, focuses on the likelihood of obtaining each posterior regardless of states rather than obtaining each act in each given state.

Also, a similar way of modeling was used first by Matějka and McKay (2013) who build a connection between rational inattention and multinominal logit models. In specific, they fix a *finite* set of acts and use the *mutual information* as the cost of information processing which together guarantee a finite number of acts to be chosen and thereby a finite number of corresponding posteriors to be considered.[3] In fact, their setting allows them to consider a discrete optimization problem and apply the Kuhn-Tucker conditions on it to solve for the optimal acts.

[3]Mutual information function is a measure of information commonly applied in the literature which is linear in information channels. See footnote 11 for the definition of mutual information.

In the current work, however, a more general setting -an arbitrary choice set, a more general type of cost function- is considered which requires a supporting hyperplane argument for the solutions of an optimal channel as shown in Theorem 5. This result extends on a solution method used in statistics literature first given by Chernoff and Reiter (1954) who investigate the optimal solutions of a bioassay problem.[4]

Finally, upon the completion of this work I got aware of a closely related study by Caplin and Dean (2013). In a similar choice setting to the one in this work, Caplin and Dean (2013) investigate the behavioral implications of rational inattention when the cost function is taken as the mutual information. In characterizing the solutions of optimal acts, they use a similar hyperplane argument to the one given in Theorem 5. The main difference between their solution result and Theorem 5 is that they only allow information channels with a finite support whereas this work considers any feasible channel and shows that an optimal channel with a finite support is always attainable when the state space is finite. In fact, it is possible also to find optimal channels with a support consisting of a number posteriors more than the number of states (see Remark 8 in section 3.2.2).

3.2 Choice Problem

In this section, I first introduce the choice problem of a Bayesian DM for whom information processing is costly and then I give a characterization for an optimal

[4]For the statement of the bioassay problem and its solutions, see Chernoff and Reiter (1954), Isaacson and Rubin (1954), Rustaghi (1957) and the related literature therein.

choice of information when the cost function is linear.[5]

3.2.1 Description of the Model

Let $\Omega = \{\omega_1, ..., \omega_n\}$ be a *finite* state space for some $n \in \mathbb{N}_+$ representing the uncertainty faced by the DM. Let $P = \Delta(\Omega)$ denote the set of probability measures over state space Ω. The DM's prior belief is denoted by $\bar{p} \in P$ assumed to have a full support. There is a set of *consequences*, denoted by X, from which the DM obtains utility according to some real-valued function $u : X \to \mathbb{R}$.

Let $\mathcal{F}$ denote the set of all *acts* (or actions), that is, the set of all functions f mapping Ω into X. As a Bayesian, the DM values each act $f \in \mathcal{F}$ for each given posterior $p \in P$ according to the standard expected utility formula:

$$U_p(f) = \int_\Omega u(f(\omega))\, p(d\omega) \tag{3.1}$$

A subset of acts $A \subset F$, called an *admissible choice set*, satisfying for each posterior $p \in P$ $\arg\max_{f \in A} U_p(f) \neq \emptyset$ is assumed to be available to the DM from which she needs to make a choice of an act.

It is assumed that the DM has access to information -via various sources- about the true state $\omega_i \in \Omega$ which can be used to make better choices. Formally, a standard approach in modeling information sources in terms of a distribution over posteriors, called *information channels,* will be used [See (Gollier, 2004, Chapter 24)].

[5]The notations and definitions given in this section mainly follow de Oliveira et al. (2013). For completeness of the discussions I provide them also here.

Information Channels

Let the set of probability measures over P be denoted by $\Delta(P)$ which is convex and compact in the topology of convergence in distribution.[6]

Definition 6 *An* (information) channel *is a probability measure* $\pi \in \Delta(P)$ *such that*

$$\bar{p}(\omega) = \int_P p(\omega)\, \pi(dp) \quad \forall \omega \in \Omega.$$

Each channel specifies a set of posteriors in P with corresponding probabilities consistent with updating the prior $\bar{p} \in P$ via Bayes rule.[7] Let $\Pi(\bar{p})$ denote the set of channels which is a convex and closed subset of $\Delta(P)$.

The notion to be used to compare channels in terms of their informativeness is the [Blackwell (1953)] order.

The Blackwell Order

Definition 7 *Let* π *and* ρ *be a pair of channels. Then* π *is called to be* more informative *than* ρ*, denoted* $\pi \trianglerighteq \rho$*, if*

$$\int_P \varphi(p)\, \pi(dp) \geq \int_P \varphi(p)\, \rho(dp)$$

for each convex function $\varphi : P \to \mathbb{R}$.

The Blackwell order intuitively requires a unanimous agreement on the comparative *usefulness* of channels. Given a choice set A, for instance, any Bayesian

[6]Convexity of $\Delta(P)$ is clear. See Aliprantis and Border (2006), p.505-506 for the compactness of it.

[7]Indeed a channel is a summary of information processing activities of a DM focusing on the posteriors.

decision maker should obtain more expected utility with channel π than channel ρ if π is *more informative* than ρ.[8]

Ideally, the DM would like to improve her choices by using the most informative channels. Following Sims (2003), however, the DM is assumed to have rationally inattentive preferences towards information. That is, the DM incurs a cost of processing information possibly due to her scarce attention. I parametrize the attention scarcity of the DM by a cost function defined as follows:[9]

Information Cost Function

Definition 8 *Given a prior $\bar{p} \in P$, a function $c : \Pi(\bar{p}) \to \mathbb{R}_+$ is called an* information cost function *if it is lower semi-continuous and satisfies the following three properties:*

(i) *Grounded*: $c(\pi) = 0$ if $\pi(\bar{p}) = 1$.

(ii) *Convex*: $c(\alpha\pi + (1-\alpha)\rho) \leq \alpha c(\pi) + (1-\alpha)c(\rho)$ if $\pi, \rho \in \Pi(\bar{p})$ and $\alpha \in [0, 1]$.

(iii) *Monotone*: $c(\pi) \geq c(\rho)$ if $\pi, \rho \in \Pi(\bar{p})$ such that $\pi \trianglerighteq \rho$.

Lower semi-continuity and convexity are standard conditions of a cost function useful for optimization purposes.[10] Groundedness is a normalization which invokes the intuitive condition that if a channel is uninformative with respect to a prior, then it is worth no attention which of course should be free. On the other

[8] For equivalent definitions of Blackwell order, see Kihlstrom (1984), pg.13-31.

[9] For an axiomatic foundation of rationally inattentive preferences and for discussions of general properties of an information cost function, see de Oliveira et al. (2013).

[10] Indeed, it follows from Theorem 5.43 in Aliprantis and Border (2006) that an information cost function c is continuous on each open convex subset of $\Pi(\bar{p})$.

hand, monotonicity ensures that more information requires more attention and paying more attention should be more costly.

In light of above notions, I can now summarize the general choice problem of interest. To ease the notation, let $\varphi_A^u(p)$ denote the utility payoff of the best act in A when the posterior is $p \in P$. That is:

$$\varphi_A^u(p) = \max_{f \in A} U_p(f), \quad \forall p \in P \tag{3.2}$$

Choice Problem Under Rational Inattention

Definition 9 *A* choice problem *consists of a tuple* C=$(A, u, \bar{p}, c)$ *such that the DM needs to solve the following optimization problem:*

$$\max_{\pi \in \Pi(\bar{p})} \left[\int_P \varphi_A^u(p)\pi(dp) - c(\pi) \right] \tag{3.3}$$

where $u : X \to \mathbb{R}$ is a utility function, $\bar{p} \in P$ is a prior belief with full support, $c : \Pi(\bar{p}) \to \mathbb{R}_+$ is an information cost function and $A \subset \mathcal{F}$ is an admissible choice set.

In (3.3), the function $\varphi_A^u(p)$ gives the value of posterior $p \in P$ when the choice set is A. The expression within square brackets then corresponds to the expected payoff of a channel $\pi \in \Pi(\bar{p})$ less the cost of it which is equal to the net value of channel π. As $\varphi_A^u(p)$ is convex, a more informative channel has a higher expected payoff while it also costs more. Taking this into account, the outer maximization finally corresponds to the optimal choice of a channel with the highest value under the choice set A.

From an ex-ante point of view, clearly the model in (3.3) leads to a stochastic choice of an act. This is due to the fact that the DM rationally chooses an act only after the realization of a posterior $p \in P$ which in turn happens randomly governed by a chosen channel. Moreover, it is clear that solutions of both maximizations in (3.2) and (3.3) depend on each other. In other words, what you learn affects what actions you would take and what actions you could take affects what you would learn. Thus, finding a solution to the problem in (3.3) could be quite hard in general and for the sake of tractability, one might need to put some restrictions on the above model.

As I will show in what follows, however, one does not need any restrictions on the type of available information when a linear cost function -such as the mutual information function- is used. Before proceeding to the next section, I now first confirm the existence of an optimal channel for the model in (3.3) showing that the choice problem in definition 9 is well-defined.

Proposition 8 *There exists an optimal channel $\pi^* \in \Pi(\bar{p})$ for a given choice problem C in (3.3). Moreover, when the cost function c is strictly convex, the optimal channel has to be unique.*

3.2.2 Characterization of a Solution

I now consider the optimal solutions of a channel $\pi^* \in \Pi(\bar{p})$ for the model in (3.3) when the cost function is linear in channels; that is, when the cost function $c : \Pi(\bar{p}) \to \mathbb{R}_+$ can be written as:

$$c(\pi) = \int_P \hat{c}(p)\, \pi(dp) \tag{3.4}$$

for some convex *posterior cost* function $\hat{c} : P \to \mathbb{R}_+$ such that $\hat{c}(\bar{p}) = 0$.[11]

It will be shown that one can find in this case a finite set of posteriors which would together form the support of an optimal channel. Before proceeding to this characterization result, some preliminary discussions would be quite useful to put. To this regard, let C be a choice problem with a cost function c satisfying condition (3.4) for some posterior cost $\hat{c} : P \to \mathbb{R}_+$ and let $V_A^{u,\hat{c}} : P \to \mathbb{R}$ denote the net-value of each posterior defined as follows:

$$V_A^{u,\hat{c}}(p) = \varphi_A^u(p) - \hat{c}(p) \tag{3.5}$$

Then the problem of the DM in (3.3) can be equivalently written as:

$$\max_{\pi \in \Pi(\bar{p})} \left[\int_P V_A^{u,\hat{c}}(p)\, \pi(dp) \right] \tag{3.6}$$

Remark 6 *The objective function in (3.6) is linear in π and therefore it is continuous. Then it immediately follows from* Bauer's maximum principle [12] *that there must exist a maximizer π^* for the problem in (3.6), possibly non-unique, which is an* extreme *point of the set* $\Pi(\bar{p})$*; that is, a point which can not be written as a strict convex combination of two distinct points in* $\Pi(\bar{p})$*. But then, it is easy to see that by* Carathodory's theorem *any such extreme point of* $\Pi(\bar{p})$ *must have a support with at most $n = |\Omega|$ posteriors in it.*

Note that the linearity of the objective function in π is essential to be able to ap-

[11]Note that the cost function c in (3.4) is well-defined in the sense that it satisfies all the properties required of an information cost function given in definition 8. Indeed, many applications of rational inattention in the literature use mutual information as the cost function which is separable in the sense of (3.4). More specifically, for the mutual information function $\hat{c}(p)$ is taken as the relative entropy function $R(p||\bar{p})$. See Cover and Thomas (2006).

[12]See Aliprantis and Border (2006), p.298-299 for the Bauer maximum principle.

ply the Bauer maximum principle in Remark 6. On the other hand, the current setting allows one to say more about optimal solutions of the problem in (3.6) by appealing to a supporting hyperplane argument.

To see this, let $L : \Delta(P) \to \mathbb{R}^n$ be a function defined for all $\pi \in \Delta(P)$ as:

$$L(\pi) = \left(\int_P p(\omega_1)\, \pi(dp), ..., \int_P p(\omega_{n-1})\, \pi(dp), \int_P V_A^{u,\hat{c}}(p)\, \pi(dp) \right) \tag{3.7}$$

Notice that the set $S = L(\Delta(P))$ is a convex subset of $\mathbb{R}^n$ since $\Delta(P)$ is convex and *linearity* of L preserves convexity. In fact, S is the convex hull of the graph of function $V_A^{u,\hat{c}}$ (see Figure 3.2 for an illustration). Moreover, for any optimal channel $\pi^* \in \Pi(\bar{p})$, $L(\pi^*)$ is a *boundary point* of S as its last coordinate is maximized while other coordinates are fixed. Therefore, there must be some hyperplane H associated with a vector $\lambda \in \mathbb{R}^{n-1}$ supporting S at $L(\pi^*)$ such that

$$L(\pi) \cdot (\lambda, 1) \leq L(\pi^*) \cdot (\lambda, 1), \quad \forall \pi \in \Delta(P) \tag{3.8}$$

But then, since H has an $n-1$ dimension, $L(\pi^*)$ can be written as a convex combination of at most n extreme points of S belonging to H. These extreme points, on the other hand, necessarily must be of the form $L(\pi_q)$ for some $q \in P$ where $\pi_q \in \Delta(P)$ denotes a point mass distribution defined as:

$$\pi_q(p) = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{if } p \neq q \end{cases} \tag{3.9}$$

I now state the above result formally. Let $\mathrm{supp}(\pi)$ denote the support of a channel $\pi \in \Pi(\bar{p})$ and $\pi_p \in \Delta(P)$ denote the point mass distribution for some $p \in P$ as defined in (3.9).

Theorem 5 *Let C be a choice problem with a linear cost function c. Then, there exists an optimal channel $\pi^* \in \Pi(\bar{p})$ with $|\operatorname{supp}(\pi^*)| \leq n$ such that for all $q \in \operatorname{supp}(\pi^*)$*

(i) $\pi_q \in \arg\max_{\pi \in \Delta(P)} [L(\pi) \cdot (\lambda, 1)]$ for some $\lambda \in \mathbb{R}^{n-1}$ and

(ii) $L(\pi_q)$ is an extreme point of the set $S = L(\Delta(P))$

where $L : \Delta(P) \to \mathbb{R}^n$ is defined as in (3.7).

A few remarks are in order.[13]

Remark 7 *Clearly, condition (i) in Theorem 5 is satisfied by* any *optimal channel $\pi^* \in \Pi(\bar{p})$ and therefore it holds for all $p \in \operatorname{supp}(\pi^*)$ for any optimal channel $\pi^* \in \Pi(\bar{p})$. Otherwise, π^* could not satisfy condition (i).*

Remark 8 *In general, one can find an optimal channel $\pi^* \in \Pi(\bar{p})$ such that $|\operatorname{supp}(\pi^*)| > n$ or condition (ii) fails for some $p \in \operatorname{supp}(\pi^*)$. In contrast, when condition (ii) is satisfied by an optimal channel, then the unique optimal channel must be $\pi_{\bar{p}}$.*

It follows from Remark 7 that finding an optimal channel is a two-step process. First, one needs to obtain a set of posteriors $q \in P$ satisfying condition (i) in Theorem 5 for some vector $\lambda \in \mathbb{R}^{n-1}$ and then find the appropriate weights $\pi^*(q)$ on each such q so that $\pi^* \in \Pi(\bar{p})$. Moreover, Theorem 5 guarantees that one can find a set of such posteriors with at most n elements in it.

In the next section, I will apply this result in a buyer-seller problem to solve for the buyer's unique optimal channel.

[13] In deriving Theorem 5, *any* choice set A is permitted as long as the choice problem C is well-defined. Matějka and McKay (2013), on the other hand, use a finite choice set with the mutual information as the cost of information while allowing for an arbitrary state space. In this regard, our works can be seen complementary.

3.3 Application: Pricing a Risky Project

In this section, I investigate a model of pricing a risky project where the seller of the project faces with a potential buyer who is rationally inattentive. The object of interest is to understand how rational inattention affects optimal pricing. To see this effect, I will first analyze the buyer's problem to derive her optimal *behavior* against each offer price and then show how the buyer's cost of information processing changes the seller's optimal prices. I now first describe the pricing problem and then analyze it in Section 3.2.

3.3.1 Description of the Pricing Problem

I consider a principal-agent setting where a risk-neutral seller s (the principal) owns a risky project and wants to sell (or rent) it to a risk-neutral buyer b (the agent) with an initial wealth of $W \in \mathbb{R}$.[14] The outcome of the project depends on the state of the world and for simplicity, I shall assume that the set of states Ω has only two elements ω_L and ω_H denoting respectively the low and the high states. Low and high state outcomes of the project are denoted by $L \in \mathbb{R}$ and $H \in \mathbb{R}$ respectively and the cost of implementing the project by $C \in \mathbb{R}_+$which are assumed to satisfy $L < C < H < W$.[15]

Let $P = \Delta(\Omega)$ denote the set of probability measures over Ω and let p_L and

[14]For instance, the risky project could be seen as a house that a landlord wants to lease or a land that the owner wants to rent it to a sharecropper. It could be seen as a patent that an entrepreneur wants to sell, a product that the customer is uncertain about the quality of it or an applicant whose fit in the job is uncertain. In fact, one can extend the list of examples of this model quite substantially.

[15]The assumption on the buyer's wealth W is placed to abstract from any wealth effects. Also, the seller is assumed to have not enough wealth or simply have no interest to take up the project.

p_H denote in short the probabilities $p(\omega_L)$ and $p(\omega_H)$ respectively for any $p \in P$. The buyer is assumed to have a prior belief $\bar{p} \in P$ such that $\bar{p}_L, \bar{p}_H \in (0,1)$.[16] I also assume that the expected payoff of the project $H - \bar{p}_L(H-L)$, denoted by E, is strictly greater than the execution cost C; that is, $E > C$. Moreover, the buyer is rationally inattentive with a linear cost function $c_\theta : \Pi(\bar{p}) \to \mathbb{R}_+$ defined for all $\pi \in \Pi(\bar{p})$ as:

$$c_\theta(\pi) = \theta \int_P \hat{c}(p)\, \pi(dp) \tag{3.10}$$

where $\theta > 0$ is a cost parameter measuring the *degree of attentiveness* and

$$\hat{c}(p) = (\bar{p}_L - p_L)^2 = \int_\Omega (\bar{p}_\omega - p_\omega)^2 \, p(d\omega)$$

for all $p \in P$. Note that the cost function c_θ in (3.10) satisfies the properties required in definition (8).[17] Also note that as parameter θ gets bigger, information processing becomes more costly or in other words the buyer becomes less attentive.

Finally, it is assumed that all parameters of the model is common knowledge. That is to say, the seller knows what the buyer's prior $\bar{p}$ is, what her information cost function c_θ is and both the seller and the buyer agree on all other parameters of the model.

[16] It is assumed for simplicity that the seller also has the same prior to abstract from any information asymmetry and therefore any strategic interaction between the buyer and the seller.

[17] This specific choice of the cost function is made for brevity. The results obtained in this work can also be derived with more common cost functions such as the mutual information.

3.3.2 Analysis of the Pricing Problem

The order of decisions is as follows: First the seller decides on an offer price $R \geq 0$ to ask for the risky project to the buyer. Facing with a price R, the buyer then decides whether to obtain information or not before taking any action: either accepting or rejecting the offer. If the buyer decides to accept the offer, she pays the seller the offer price R, incurs the cost of executing the project C and finally obtains a payoff of L or H depending on the state of the world. If, on the other hand, she decides not to take the offer, then no transaction occurs between the buyer and the seller. In this case, the seller does not obtain any revenue and the buyer keeps her wealth W.[18] I now formally analyze the buyer and the seller problem starting from the former.

Buyer's Problem

As the buyer is risk neutral, I first normalize her utility function such that $u_b(x) = x$ for any $x \in \mathbb{R}$. Given an offer price R, the buyer first needs to decide whether to obtain information or not before making a choice of an act: *accepting* or *declining* the offer denoted by a_R and d respectively. If she chooses act a_R, her payoff would be $(L - R) + (W - C)$ when the state is low and $(H - R) + (W - C)$ when it is high. On the other hand, by choosing act d she can only obtain a constant payoff at her initial wealth W. Thus, the buyer's choice problem can be summarized with the tuple $C_{R,\theta} = (A_R, u_b, \bar{p}, c_\theta)$ where

$$A_R = \{a_R, d\} \tag{3.11}$$

[18]Note that the seller gives the buyer an option of accepting or declining the offer with or without obtaining information. An alternative pricing scheme, in fact, could also value this option. See section 3.4 for more on this.

It is clear that for a given posterior $p \in P$, the buyer will accept an offer R if and only if its expected benefit, $E_p = [H - p_L(H - L)]$ is more than its cost $R + C$. That is, the offer will be accepted when $p \in P$ is such that $p_L \leq \frac{H-R-C}{H-L}$. Thus for a given offer R, the value of a posterior $p \in P$ for the buyer is:

$$\varphi_{A_R}(p) = \max_{f \in A_R} \left(\int_\Omega f(\omega)\, p(d\omega) \right) = \begin{cases} W + E_p - (R + C) & \text{if } p_L \leq \frac{H-R-C}{H-L} \\ W & \text{if } p_L > \frac{H-R-C}{H-L} \end{cases} \tag{3.12}$$

Now, the buyer's objective is to use a channel for the problem $C_{R,\theta}$ optimizing the expression in (3.6). That is, to find an optimal channel solving the following problem:

$$\max_{\pi \in \Pi(\bar{p})} \left[\int_P V_{A_R}^{\theta,\hat{c}}(p) \pi(dp) \right] \tag{3.13}$$

where for all $p \in P$

$$V_{A_R}^{\theta,\hat{c}}(p) = \varphi_{A_R}(p) - \theta \hat{c}(p)$$

denotes the net-value of each posterior. Figure 3.1 below illustrates the value $\varphi_{A_R}(p)$, cost $\hat{c}(p)$, and net-value $V_{A_R}^{\theta,\hat{c}}(p)$ of each posterior $p \in P$ for some given offer R and parameter θ. Figure 3.2 then shows the convex hull of the graph of function $V_{A_R}^{\theta,\hat{c}}(p)$ and the optimal posteriors for the problem in (3.13).

By Proposition 8 an optimal channel $\pi^* \in \Pi(\bar{p})$ exists and by Theorem 5 it can be uniquely defined with a support of at most two points. Indeed, Proposition 9 below lists the support of an optimal channel for all pairs of admissible offer and cost parameter (R, θ). Before stating this result, some new notation would

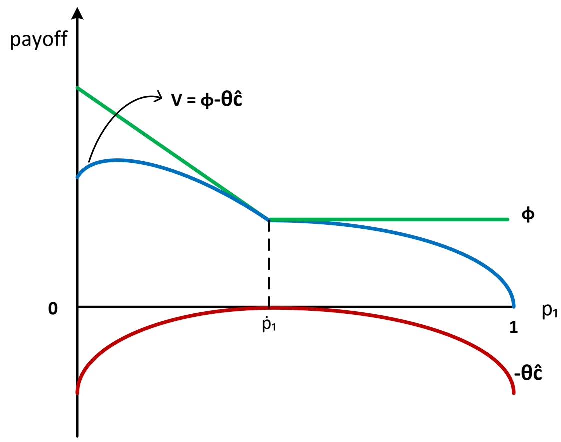

Figure 3.1: A sketch of value, cost and net-value of posteriors.

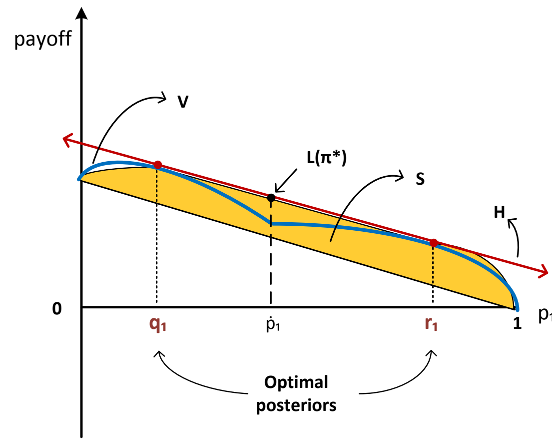

Figure 3.2: Convex hull of posterior/net-values and optimal posteriors .

be useful to introduce. Let $\theta_0(R) = 0$ for all $R \in [H - C, \infty)$ and let

$$\theta_1(R) = \begin{cases} \frac{(H-L)^2}{4(C+R-L)} & \text{if } R \in [0, \max\{\frac{H+L}{2} - C, 0\}) \\ H - C - R & \text{if } R \in [\max\{\frac{H+L}{2} - C, 0\}, H - C) \end{cases}$$

$$\theta_2(R) = \begin{cases} \frac{C+R-L}{(\bar{p}_H)^2} & \text{if } R \in [0, \max\{\frac{E+L}{2} - C, 0\}) \\ \frac{(H-L)^2}{4(E-C-R)} & \text{if } R \in [\max\{\frac{E+L}{2} - C, 0\}, E - C) \end{cases}$$

$$\theta_3(R) = \begin{cases} \frac{(H-L)^2}{4(C+R-E)} & \text{if } R \in (E - C, \frac{E+H}{2} - C] \\ \frac{H-C-R}{(\bar{p}_L)^2} & \text{if } R \in [\frac{E+H}{2} - C, H - C) \end{cases} \tag{3.14}$$

$$\theta_4(R) = \begin{cases} C + R - L & \text{if } R \in [0, \max\{\frac{H+L}{2} - C, 0\}) \\ \frac{(H-L)^2}{4(H-C-R)} & \text{if } R \in [\max\{\frac{H+L}{2} - C, 0\}, H - C) \end{cases}$$

And finally let

$$
\begin{aligned}
I_1 &= \operatorname{epi}\theta_2 & I_4 &= \operatorname{epi}\theta_1 \cap \operatorname{hyp}_S \theta_3 \cap \operatorname{hyp}\theta_4 \\
I_2 &= \operatorname{epi}\theta_1 \cap [\operatorname{hyp}_S \theta_2 \cup \operatorname{hyp}_S \theta_3] \cap \operatorname{epi}\theta_4 & I_5 &= \operatorname{hyp}\theta_1 \cap \operatorname{hyp}\theta_4 \\
I_3 &= \operatorname{hyp}\theta_1 \cap \operatorname{hyp}_S \theta_2 \cap \operatorname{epi}\theta_4 & I_6 &= \operatorname{epi}\theta_3 \cup \operatorname{epi}_S \theta_5
\end{aligned}
\tag{3.15}
$$

where epi, epi_S, hyp and hyp_S denote the epigraph, strict epigraph, hypograph and strict hypograph of a function respectively. Note that the sets, I_1 to I_6 defined in (3.15) partition the region $[0, \infty) \times (0, \infty)$ into six sub-regions as depicted in Figure 3.3 below.

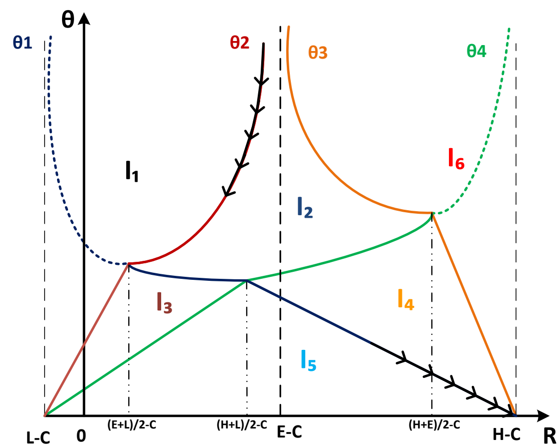

Figure 3.3: Sub-regions and motion of optimal price.

I am now ready to state the proposition giving a full characterization of the unique optimal channel for the problem $C_{R,\theta}$ with its support.

Proposition 9 *Let $C_{R,\theta}$ be the buyer's problem for some given cost parameter $\theta > 0$ and offer price $R \geq 0$. Then the unique optimal channel $\pi^*_{R,\theta} \in \Pi(\bar{p})$ has its support as*

$\{p^{R,\theta}, q^{R,\theta}\} \subset P$ *where* $p_L^{R,\theta}$ *and* $q_L^{R,\theta} \in [0,1]$ *are given as:*

$$
(p_L^{R,\theta}, q_L^{R,\theta}) = \begin{cases} (\bar{p}_L, \bar{p}_L) & \textit{if } (R,\theta) \in I_1 \cup I_6 \\ \left(\frac{H-C-R}{H-L} - \frac{H-L}{4\theta}, \frac{H-C-R}{H-L} + \frac{H-L}{4\theta}\right) & \textit{if } (R,\theta) \in I_2 \\ \left(1 - \sqrt{\frac{C+R-L}{\theta}}, 1\right) & \textit{if } (R,\theta) \in I_3 \\ \left(0, \sqrt{\frac{H-C-R}{\theta}}\right) & \textit{if } (R,\theta) \in I_4 \\ (0,1) & \textit{if } (R,\theta) \in I_5 \end{cases}
$$

The support given in Proposition 9 for each (R,θ) uniquely identifies an optimal channel due to the fact that the support has only at most two posteriors in it and therefore can be mixed in only one way to obtain the prior. These unique weights on each posterior then necessarily defines the optimal channel.

Now, if the support has only the prior $\bar{p}$ as its element, the buyer does not process any information and therefore deterministically accepts or rejects the offer depending on how good the offer is. In fact it follows from (3.12) that when $(R,\theta) \in I_1$, that is when the offer is relatively cheap enough, the buyer accepts the offer as she has $\bar{p}_L \le \frac{H-C-R}{H-L}$ in this case and in contrast when $(R,\theta) \in I_6$, that is when the offer is relatively expensive enough, she rejects the offer since then $\bar{p}_L > \frac{H-C-R}{H-L}$.

On the other hand, when the support has two distinct points $p^{R,\theta}$and $q^{R,\theta}$ in it, the buyer processes some information and accepts or rejects the offer depending on the stochastic realization of a posterior. Indeed, she accepts the offer when the realized posterior is $p^{R,\theta}$ since $p_L^{R,\theta} \leq \frac{H-C-R}{H-L}$ when $(R,\theta) \notin I_1 \cup I_6$ and rejects the offer when $q^{R,\theta}$ is realized since $q_L^{R,\theta} > \frac{H-C-R}{H-L}$ when $(R,\theta) \notin I_1 \cup I_6$.

The following proposition summarizes above discussion and gives the acceptance likelihood $\pi_a(R,\theta) \in [0,1]$ of an offer $R \geq 0$ when the buyer has a cost parameter $\theta > 0$. I will particularly use this result when the seller's problem is considered in the next section.

Proposition 10 *Let $C_{R,\theta}$ be the buyer's problem for some given cost parameter $\theta > 0$ and offer price $R \geq 0$ and let $\{p^{R,\theta}, q^{R,\theta}\}$ be the support of the unique optimal channel $\pi^*_{R,\theta}$. Then the probability of accepting the offer $\pi_a(R,\theta) \in [0,1]$ can be given as follows:*

$$\pi_a(R,\theta) = \begin{cases} 1 & \text{if } (R,\theta) \in I_1 \\ \frac{q_L^{R,\theta} - \bar{p}_L}{q_L^{R,\theta} - p_L^{R,\theta}} \in (0,1) & \text{if } (R,\theta) \in I_2 \cup I_3 \cup I_4 \cup I_5 \\ 0 & \text{if } (R,\theta) \in I_6 \end{cases} \tag{3.16}$$

Remark 9 *For a given θ, the acceptance likelihood $\pi_a(R,\theta)$ is non-increasing in R. That is, the buyer becomes less likely to accept an offer as the price of the offer gets higher. This intuitive result holds due to the fact that the buyer would like to accept an offer only when she is certain enough that the project brings her a high payoff. That is, only when $p_L \leq \frac{H-R-C}{H-L}$ as given in (3.12). But as R gets bigger, the right hand side of this inequality becomes smaller and thus the likelihood of the posterior at which the offer is accepted, $p^{R,\theta}$, becomes lower.*

As just noted in Remark 9 above, the seller may obtain more revenue by risking the acceptance of the offer. Thus, for the seller there is a certain tradeoff between the revenue of an offer and the likelihood of its acceptance. In the following section, how this tradeoff is affected as the buyer's cost of processing information changes will be investigated in more detail.

Seller's Problem

As is the seller also risk-neutral, I first normalize and take his payoff function as $u_s(x) = x$ for any $x \in \mathbb{R}$. Now, given a cost parameter $\theta > 0$, the objective of the seller is to set a price $R \geq 0$ giving him the highest expected payoff. That is, he needs to maximize his expected revenue given as:

$$\max_{R \geq 0} \pi_a(R, \theta) \times R \tag{3.17}$$

where $\pi_a(R, \theta)$ is the probability of the buyer accepting the offer defined as in (3.16).

Remark 10 *By using Propositions 9 and 10, one can easily verify that the acceptance likelihood of an offer $\pi_a(R, \theta)$ is continuous in R for any given θ and so is the objective function in (3.17). Moreover, it is clear that an optimal offer price R can not be higher than $H - C$ since otherwise the seller can not obtain any revenue whereas he could earn some for an offer price $R \in [0, H - C]$. Thus, one can restrict the domain of the maximization in (3.17) to the interval $[0, H - C]$ implying that the seller's problem in (3.17) has a solution and so it is well-defined.*

Now, the subject of interest is to understand how these optimal prices $R^*(\theta)$ are affected by the changes in the buyer's cost parameter θ. Should the seller increase the ask price R as the buyer becomes more attentive or should he decrease it? The following proposition shows that, perhaps surprisingly, the optimal price $R^*(\theta)$ should be non-monotonic in the cost parameter θ.

Proposition 11 *Let the seller's problem be given as in (3.17) for some cost parameter* $\theta > 0$*. Then, there exist some* $\theta' > \theta'' > 0$ *such that the seller's optimal price* $R^*(\theta)$ *satisfies:*

$$R^*(\theta) = \begin{cases} E - C - \frac{(H-L)^2}{4\theta} & \text{when } \theta > \theta' \\ \\ H - C - \theta & \text{when } \theta < \theta'' \end{cases}$$

Note that the optimal price $R^*(\theta)$ is monotonically decreasing on the interval $(0, \theta'')$ and increasing on $\theta \in (\theta', \infty)$ as also depicted in Figure 3.3 above. That is to say, when the buyer is sufficiently *attentive* ($\theta < \theta''$), it is optimal for the seller to increase the price as θ gets smaller. On the other hand, when the buyer is sufficiently *inattentive* ($\theta > \theta'$), the seller finds it optimal to decrease the price as θ gets smaller. The reason for this behavior is that when $\theta < \theta''$, benefit of increasing the price is more than cost of risking the offer's acceptance . In contrast, when $\theta > \theta'$, benefit of not risking the offer's acceptance outweighs cost of loosing payoff by decreasing the price .

It is worth noting that as θ approaches to ∞, the optimal price $R^*(\theta)$ increases towards $E - C$ which is the optimal price to offer when the buyer can not obtain information about the true state. Also, when θ tends to 0, the optimal price $R^*(\theta)$ increases towards $H - C$ which is the optimal price to ask for if the buyer can always perfectly identify the true state.

Finally, note that Proposition 11 gives a partial characterization for the seller's optimal price as it is silent on the nature of $R^*(\theta)$ over the interval $[\theta'', \theta']$. This is due to the fact that many possibilities arise and to focus on a particular behavior one needs to invoke further assumptions on the parameters of the model. This task will be taken up in future research as it is discussed in more detail in the next section.

3.4 Discussion and Concluding Remarks

Rational inattention theory has found use in my different choice settings most notably on macroeconomic and finance problems. To this date, however, its applicability has been relied on strong assumptions on the parameters of respective models, especially on the type of available information. I believe the results obtained in section 2 of this work might shed some light on the theory and applications of rational inattention which hopefully would flourish more research in particular stemming from microeconomics based questions.

To this regard, a simple buyer-seller problem is considered in section 3 to demonstrate the use of Theorem 5. Moreover, an interesting feature of rational inattention on the optimal pricing of the risky project is shown. Specifically, I found that the optimal price $R^*(\theta)$ should be non-monotonic in the cost parameter θ which measures the attentiveness of the buyer. In this simple model, I treated θ as an exogenous parameter as part of the definition of the buyer about which the seller is perfectly knowledgeable. Indeed, it is straightforward to extend the model to the case where the seller has some uncertainty about the cost parameter θ.

On the other hand, one might also see parameter θ as a description of the economic stage where the buyer and the seller are in interaction. In specific, θ might be viewed as measuring how *hard* or *easy* to transmit information from a source to the buyer. For instance, the number of the Internet providers, wireless phone companies, TV channels and such might determine the degree of this parameter. An interesting exercise in this setting then would be to look for the optimal level of θ if the goal is to achieve a maximum social welfare according to some welfare function. Or parameter θ could be taken as a *costly* advertisement intensity controlled by the seller affecting how easy or hard to obtain information about the risky project. Then, in this setting one might look for the seller's optimal levels of price R^* and parameter θ^*.

Finally, in this simple model it is implicitly assumed that the seller forgoes the value of providing an option to the seller. That is, the option of being able to study and accept the offer conditional on the posterior knowledge. In an alternative setting, the seller could offer a scheme of prices (r, R) where $r \geq 0$ denotes the price of being able to buy the project later at a price not more than $R \geq 0$. It is easy to see that when the buyer is sufficiently inattentive, then it would not matter for the seller whether to put an option price or not. However, this might not necessarily be the case when the buyer is sufficiently attentive as then she accepts the offer in a stochastic manner. It would be interesting to see in that case whether a unique optimal price scheme can be obtained or not. And if the answer is yes, then one can ask for the effects of parameter θ on these prices. If not, then also it would be interesting to see how these prices are related to each other as the parameter θ varies. I postpone the quest for answers of these and related questions to future research.

3.5 Appendix

Proof of Proposition 8:

Consider the expression in (3.3) for a given choice problem C. Let $\varphi : \Pi(\bar{p}) \to \mathbb{R}$ be defined for all $\pi \in \Pi(\bar{p})$ as:

$$\varphi(\pi) = \int_P \varphi_A^u(p)\pi(dp)$$

Clearly, the function φ is linear in π and therefore continuous. Also, it is given that the cost function $c : \Pi(\bar{p}) \to \mathbb{R}$ in (3.3) is weakly lower semi-continuous. Thus, the function $[\varphi - c] : \Pi(\bar{p}) \to \mathbb{R}$ defined as $[\varphi - c](\pi) = \varphi(\pi) - c(\pi)$ for all $\pi \in \Pi(\bar{p})$ is weakly upper semi-continuous. Now, the problem in (3.3) can be equivalently written as:

$$\max_{\pi \in \Pi(\bar{p})} [\varphi - c](\pi) \tag{3.18}$$

Since the objective function in (3.18) is upper semi-continuous and its domain $\Pi(\bar{p})$ is weakly compact, a maximum must be achieved.[19] Moreover, when the cost function c is strictly convex, then the objective function $[\varphi - c]$ is strictly concave and so the optimal channel must be unique.

Proof of Theorem 5:

Let C be a choice problem with a linear cost function c and let $L : \Delta(P) \to \mathbb{R}^n$ be a function defined for all $\pi \in \Delta(P)$ as in (3.7). Then, the set $S = L(\Delta(P))$ is a compact convex subset of $\mathbb{R}^n$ since L is a linear function and thus preserves convexity and compactness. As it is argued in the text, the point $L(\pi^*)$ is a boundary point of S for any optimal channel $\pi^* \in \Pi(\bar{p})$. Therefore, there must be some hyperplane

[19] See Aliprantis and Border Aliprantis and Border (2006), Theorem 2.43.

$H_\lambda \subset \mathbb{R}^n$ associated with a vector $\lambda \in \mathbb{R}^{n-1}$ supporting S at $L(\pi^*)$.[20] That is, $H_\lambda = \{h \in \mathbb{R}^n : h \cdot (\lambda, 1) = L(\pi^*) \cdot (\lambda, 1)\}$ for some $\lambda \in \mathbb{R}^{n-1}$ such that for any $\pi \in \Delta(P)$ the inequality in (3.8) holds. In particular, (3.8) holds with equality for any $L(\pi) \in H_\lambda \cap S$. Also note that H_λ is a closed convex set and therefore $H_\lambda \cap S$ must be a compact convex set.

But then, since $H_\lambda \cap S$ has an $n-1$ dimension, by Minkowski-Carathodory theorem $L(\pi^*)$ can be written as a convex combination of at most n extreme points of S belonging to H_λ.[21] Moreover, these extreme points necessarily must be of the form $L(\pi_q)$ for some $q \in P$ where $\pi_q \in \Delta(P)$ is a point distribution defined as in (3.9). Then, clearly conditions (i) and (ii) in Theorem 5 are satisfied for each $q \in P$.

Proof of Proposition 9:

Let $C_{R,\theta}$ be the buyer's problem defined in (3.11) for some given cost parameter $\theta > 0$ and offer price $R \geq 0$. By Proposition 8, there exists an optimal channel $\pi^*_{R,\theta} \in \Pi(\bar{p})$ solving the expression in (3.13). Moreover, by Theorem 5 optimal channels can have at most two points in their support. I now want to show that the unique optimal channel can be obtained by identifying its support.

First note that when $R \geq H - C$, for all $p \in P$

$$\int_\Omega u_b(d(\omega))\, p(d\omega) \geq \int_\Omega u_b(a_R(\omega))\, p(d\omega) \tag{3.19}$$

and moreover for some $p \in P$, the inequality in (3.19) is strict. Thus, when $(R, \theta) \in \mathrm{epi}_S\, \theta_5$ the unique optimal channel trivially has support $\mathrm{supp}(\pi^*_{R,\theta}) = \{\bar{p}\}$

[20]See Aliprantis and Border (2006) for the general statement of the supporting hyperplane theorem.

[21]See Simon Simon (2011), theorem 8.11.

where $\text{epi}_S\,\theta_5$ is defined as in (3.15). Hence, for the remainder of the proof I consider the case where $R \in [0, H - C)$.

Now define the function $V_R : P \to \mathbb{R}$ as in (3.5) for the buyer's problem $C_{R,\theta}$ and let $V : [0,1] \to \mathbb{R}$ be a function such that for all $x \in \mathbb{R}$, $V(x) = V_R((x, 1-x))$. That is,

$$V(x) = \begin{cases} (E_x - R) + (W - C) - \theta(\bar{p}_L - x)^2 & \text{if } x \leq x_R \\ W - \theta(\bar{p}_L - x)^2 & \text{if } x > x_R \end{cases}$$

where $E_x = H - x(H - L)$ and $x_R = \frac{H-C-R}{H-L}$.

Note that, $0 < x_R < 1$ by the fact that $L < C + R < H$. Moreover, it can be easily seen that $V(x)$ is piecewise differentiable over $[0,1]$ except at $x = x_R$ and is strictly concave in x over each of its sub-domains $[0, x_R)$ and $(x_R, 1]$.[22]

Now, from the proof of Theorem 5 it follows that finding an optimal channel for the problem in (3.13) is equivalent to finding a point s^* in the set $S = \{(\int_{[0,1]} x\,\pi(dx),\ \int_{[0,1]} V(x)\,\pi(dx)) \in \mathbb{R}^2\}$ with the first coordinate fixed at $\bar{p}_L$ and second coordinate maximized. Also it follows that such a point s^* is a boundary point of set S and so can be written as a convex combination of at most two extreme points of S which lie on the same line supporting S at point s^*. Therefore, there exist some $\lambda \in \mathbb{R}$ for s^* such that for all $(s_1, s_2) \in S$

$$\lambda s_1 + s_2 \leq \lambda \bar{p}_L + s_2^* \tag{3.20}$$

and if s^* is not an extreme point itself, then there must exist some $x^* \in [0, \min\{\bar{p}_L, x_R\})$ and $y^* \in (\max\{\bar{p}_L, x_R\}, 1]$ such that $(x^*, V(x^*))$ and $(y^*, V(y^*))$ are extreme points of set S and satisfy (3.20) with equality. Note that any extreme point of S must have the form as $(z, V(z))$ for some $z \in [0,1]$ and also

[22] Since V is semi-differentiable at points $x = 0$ and $x = 1$, I consider the right and left derivatives at these points respectively.

due to the strict concavity of V on its sub-domains, there can not be any $z^* \in [\min\{\bar{p}_L, x_R\}, \max\{\bar{p}_L, x_R\}]$ where $(z^*, V(z^*))$ is an extreme point of S satisfying (3.20) with equality unless s^* is an extreme point.

Now I want to find when s^* is not an extreme point and thus can be written as a convex combination of two extreme points $(x^*, V(x^*))$ and $(y^*, V(y^*))$. Each such extreme point $(z, V(z)) \in S$ must satisfy (3.20) with equality or in other words maximizes $\lambda z + V(z)$. But since V is strictly concave and differentiable respectively on the intervals $[0, \min\{\bar{p}_L, x_R\})$ and $(\max\{\bar{p}_L, x_R\}, 1]$, the following first order conditions are necessary and sufficient to find optimal $x^*$and y^* in terms of λ:

$$V'(z) + \lambda \leq 0 \quad \text{if } z = 0$$
$$V'(z) + \lambda = 0 \quad \text{if } z \neq 0 \text{ or } 1$$
$$V'(z) + \lambda \geq 0 \quad \text{if } z = 1$$

And in conjunction with the equality $V(x^*) + \lambda x^* = V(y^*) + \lambda y^*$ that they must satisfy, optimal $x^*$and y^* can be found free of λ. As noted above there are several cases to consider: (i) $x^* \in (0, \min\{\bar{p}_L, x_R\})$ and $y^* \in (\max\{\bar{p}_L, x_R\}, 1)$, (ii)$x^* \in (0, \min\{\bar{p}_L, x_R\})$ and $y^* = 1$, (iii) $x^* = 0$ and $y^* \in (\max\{\bar{p}_L, x_R\}, 1)$ and (iv) $x^* = 0$ and $y^* = 1$.

Case (i): In this case, the optimality conditions are as follows:

$$\begin{aligned} V(x^*) + \lambda x^* &= V(y^*) + \lambda y^* \\ V'(x^*) + \lambda &= 0 = V'(y^*) + \lambda \end{aligned} \tag{3.21}$$

From conditions in (3.21), I obtain:

$$\begin{aligned} x^* &= \tfrac{H-C-R}{H-L} - \tfrac{H-L}{4\theta} \\ y^* &= \tfrac{H-C-R}{H-L} + \tfrac{H-L}{4\theta} \end{aligned} \tag{3.22}$$

and by checking whether $x^* \in (0, \min\{\bar{p}_L, x_R\})$ and $y^* \in (\max\{\bar{p}_L, x_R\}, 1)$, I derive that $(R, \theta) \in I_2$ where I_2 is defined in (3.15). Therefore, when $(R, \theta) \in I_2$, $\text{supp}(\pi^*_{R,\theta}) = \{(x^*, 1 - x^*), (y^*, 1 - y^*)\}$ where x^*, y^* are given in (3.22).

Case (ii): In this case, the optimality conditions are as follows:

$$\begin{aligned} V(x^*) + \lambda x^* &= V(1) + \lambda \\ V'(x^*) + \lambda = 0 &\leq V'(1) + \lambda \end{aligned} \tag{3.23}$$

From conditions in (3.23), I obtain:

$$x^* = 1 - (\tfrac{C+R-L}{\theta})^{\frac{1}{2}} \tag{3.24}$$

and by checking the domain condition $x^* \in (0, \min\{\bar{p}_L, x_R\})$, I derive that $(R, \theta) \in I_3$ where I_3 is defined in (3.15). Therefore, when $(R, \theta) \in I_3$, the support is $\text{supp}(\pi^*_{R,\theta}) = \{(x^*, 1 - x^*), (1, 0)\}$ where x^* is given in (3.24).

Case (iii): In this case, the optimality conditions are as follows:

$$\begin{aligned} V(0) &= V(y^*) + \lambda y^* \\ V'(0) + \lambda \leq 0 &= V'(y^*) + \lambda \end{aligned} \tag{3.25}$$

From conditions in (3.25), I get:

$$y^* = (\tfrac{H-C-R}{\theta})^{\frac{1}{2}} \tag{3.26}$$

and by checking the domain condition $y^* \in (\max\{\bar{p}_L, x_R\}, 1)$, I obtain that $(R, \theta) \in I_4$ where I_4 is defined in (3.15). Thus, when $(R, \theta) \in I_4$ the support is $\text{supp}(\pi^*_{R,\theta}) = \{(0, 1), (y^*, 1 - y^*)\}$ where y^* is given in (3.26).

Case (iv): In this case, the optimality conditions are as follows:

$$\begin{aligned} V(0) &= V(1) + \lambda \\ V'(0) + \lambda \leq 0 &\leq V'(1) + \lambda \end{aligned} \tag{3.27}$$

From conditions in (3.27), I obtain that $(R, \theta) \in I_5$ where I_5 is defined in (3.15). Thus, when $(R, \theta) \in I_5$ the support is $\text{supp}(\pi^*_{R,\theta}) = \{(0, 1), (1, 0)\}$.

From above derivations, one can conclude that when $(R, \theta) \in \operatorname{epi}\theta_2 \cup \operatorname{epi}\theta_3$ which are defined in (3.15), s^* can *not* be written as a convex combination of two extreme points satisfying (3.20) and thus s^* must be equal to $(\bar{p}_L, V(\bar{p}_L))$ as an extreme point of S. That is to say, when $(R, \theta) \in \operatorname{epi}\theta_2 \cup \operatorname{epi}\theta_3$, one have $\operatorname{supp}(\pi^*_{R,\theta}) = \{\bar{p}\}$.

Proof of Proposition 10:

Let $C_{R,\theta}$ be the buyer's problem defined in (3.11) for some given cost parameter $\theta > 0$ and offer price $R \geq 0$. For any posterior $p \in P$, the buyer accepts the offer if and only if $(H - R) - p_L(H - L) + (W - C) \geq W$. That is, she chooses act a_R if and only if she has a posterior $p \in P$ satisfying

$$p_L \leq \frac{H - R - C}{H - L} = x_R \tag{3.28}$$

Now let $(R, \theta) \in I_2 \cup I_3 \cup I_4 \cup I_5$. From Proposition 9, the optimal channel $\pi^*_{R,\theta} \in \Pi(\bar{p})$ for the problem $(A_R, u_b, \bar{p}, c_\theta)$ has its support as $\operatorname{supp}(\pi^*_{R,\theta}) = \{p^{R,\theta}, q^{R,\theta}\} \subset P$ where $p^{R,\theta}_L < \min\{\bar{p}_L, x_R\}$ and $q^{R,\theta}_L > \max\{\bar{p}_L, x_R\}$. Thus, in this case the buyer accepts the offer only when she has the posterior $p^{R,\theta}$ by condition (3.28). Also from the definition of a channel, it must be the case that $\pi^*_{R,\theta}(p^{R,\theta})p^{R,\theta}_L + (1 - \pi^*_{R,\theta}(p^{R,\theta}))q^{R,\theta}_L = \bar{p}_L$. Therefore, the buyer accepts the offer with the probability given as:

$$\pi^*_{R,\theta}(p^{R,\theta}) = \frac{q^{R,\theta}_L - \bar{p}_L}{q^{R,\theta}_L - p^{R,\theta}_L} \in (0, 1)$$

Now let $(R, \theta) \in I_1$. From Proposition 9, the buyer's optimal channel has a support as $\operatorname{supp}(\pi^*_{R,\theta}) = \{\bar{p}\}$. That is, she uses only her prior in making the choice of an act. From the definition of I_1 given in (3.15), I get $R < E - C$. More explicitly, $R < H - C - \bar{p}_L(H - L)$ and so by condition (3.28) the buyer accepts the offer with probability 1.

Now let $(R, \theta) \in I_6$. By the proof of Proposition 9, the support is $\mathrm{supp}(\pi^*_{R,\theta}) = \{\bar{p}\}$ when $(R, \theta) \in \mathrm{epi}_S\, \theta_5$ and she does not accept the offer in this case. Finally, consider the case $(R, \theta) \in \mathrm{epi}\, \theta_3$. Once again, from Proposition 9 it follows that $\mathrm{supp}(\pi^*_{R,\theta}) = \{\bar{p}\}$. In this case, I have $R > E - C$ and so $\bar{p}_L > x_R$. Thus, by condition (3.28) the buyer rejects the offer with probability 1.

Proof of Proposition 11:

Let the seller's problem be defined as in (3.17) for some given cost parameter $\theta > 0$. By Remark 10, an optimal price $R^*(\theta)$ for the seller's problem must be in $[0,\ H - C]$. Moreover, the optimal price $R^*(\theta)$ can be uniquely given as (i) $R^*(\theta) = E - C - \frac{(H-L)^2}{4\theta}$ for θ sufficiently high and as (ii) $R^*(\theta) = H - C - \theta$ for θ sufficiently low. I now show the first part of the claim.

Let $\theta \in (0, \infty)$ be sufficiently high such that there exist some $R_2(\theta) \in (\max\{\frac{E+L}{2} - C, 0\},\ E - C)$ and $R_3(\theta) \in (E - C,\ \frac{E+H}{2} - C)$ satisfying:

$$\theta_2(R_2(\theta)) = \theta = \theta_3(R_3(\theta)) \tag{3.29}$$

By (3.14), there exists such a $\theta \in (0, \infty)$ satisfying (3.29). Also note that in this case, one has $([0,\ H - C] \times \{\theta\}) \cap (I_3 \cup I_4 \cup I_5) = \emptyset$ by (3.15) and therefore by Propositions 10 and 9, the acceptance probability $\pi_a(R, \theta) \in [0, 1]$ can be given as:

$$\pi_a(R, \theta) = \begin{cases} 1 & \text{if } R \in [0,\ R_2(\theta)] \\ \frac{1}{2} + \frac{2\theta}{H-L}\left(\frac{H-R-C}{H-L} - \bar{p}_L\right) & \text{if } R \in [R_2(\theta),\ R_3(\theta)] \\ 0 & \text{if } R \in [R_3(\theta),\ H - C] \end{cases}$$

Note that an optimal price $R^*(\theta)$ must necessarily be in $[R_2(\theta),\ R_3(\theta)]$. I now check that the seller's objective function $\pi_a(R, \theta) \times R$ is strictly concave over $[R_2(\theta),\ R_3(\theta)]$

and thus the optimal price must be unique. It is clear that $\pi_a(R,\theta)\times R$ is continuously differentiable in R over the interval and thus verifying the strict concavity of $\pi_a(R,\theta)\times R$ amounts to checking the following condition:

$$\frac{\partial^2\pi_a(R,\theta)}{\partial R}R+2\frac{\partial\pi_a(R,\theta)}{\partial R}<0 \tag{3.30}$$

In this case, the second order partial derivative in R vanishes and since $\frac{\partial\pi_a(R,\theta)}{\partial R}=-\frac{2\theta}{(H-L)^2}<0$, I conclude that $\pi_a(R,\theta)\times R$ is strictly concave. Now I want to see when the optimal price $R^*(\theta)$ is equal to $R_2(\theta)$ where I have $R_2(\theta)=E-C-\frac{(H-L)^2}{4\theta}$ from (3.14) . As $\pi_a(R,\theta)\times R$ is strictly concave, I have the following necessary and sufficient first order condition for the optimality of $R_2(\theta)$:

$$\frac{\partial\pi_a(R_2(\theta),\theta)}{\partial R}R_2(\theta)+\pi_a(R_2(\theta),\theta)\leq 0 \tag{3.31}$$

It can be easily shown that condition (3.31) holds if and only if $\theta\geq\frac{3(H-L)^2}{2(E-C)}$. Thus for sufficiently high θ, I have $R^*(\theta)=E-C-\frac{(H-L)^2}{4\theta}$ verifying the first part of the claim.

Now I show the second part of the above claim. For this, let θ be sufficiently low such that $\theta\leq C-L$. It can be easily verified by checking (3.15) that $([0,\ H-C]\times\{\theta\})\cap(I_2\cup I_3)=\emptyset$. Moreover, the sub-regions I_5,I_4 and I_1decomposes the interval $[0,\ H-C]$ into $[0,\ H-C-\theta)$, $[H-C-\theta,\ H-C-\theta\,\bar{p}_L^2]$ and $(H-C-\theta\,\bar{p}_L^2,H-C]$. Then by Propositions 10 and 9, the acceptance probability $\pi_a(R,\theta)\in[0,1]$ can be given as:

$$\pi_a(R,\theta)=\begin{cases}1-\bar{p}_L & \text{if } R\in[0,\ H-C-\theta)\\ 1-\bar{p}_L(\frac{\theta}{H-C-R})^{1/2} & \text{if } R\in[H-C-\theta,\ H-C-\theta\,\bar{p}_L^2]\\ 0 & \text{if } R\in(H-C-\theta\,\bar{p}_L^2,H-C]\end{cases}$$

Clearly, the optimal price $R^*(\theta)$ must be in $[H-C-\theta,\ H-C-\theta\bar{p}_L^2]$. Now I want to see when the seller's objective function $\pi_a(R,\theta)\times R$ achieves its maximum at

$H - C - \theta$. First I verify that $\pi_a(R, \theta) \times R$ is strictly concave in R on the interval $[H - C - \theta,\ H - C - \theta\bar{p}_L^2]$. To do this I need to check the condition in (3.30). It can be easily verified that condition (3.30) holds if and only if $R < 4\,(H - C)$ which is true on the interval $[H - C - \theta,\ H - C - \theta\bar{p}_L^2]$ and so $\pi_a(R, \theta) \times R$ is strictly concave. Likewise condition (3.31) above, the necessary and sufficient condition for the point $H - C - \theta$ to be optimal is:

$$\frac{\partial \pi_a((H - C - \theta), \theta)}{\partial R} \times (H - C - \theta) + \pi_a((H - C - \theta), \theta) \leq 0 \tag{3.32}$$

It can be easily shown that condition (3.32) can hold if and only if $\frac{1-\bar{p}_L}{\bar{p}_L} \leq \frac{H-C-\theta}{2\theta}$. Clearly, as θ gets smaller this inequality will hold. Thus, I conclude that for θ sufficiently low the optimal price $R^*(\theta)$ is equal to $H - C - \theta$ as desired.

Bibliography

C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis*. Springer, third edition, 2006.

F. J. Anscombe and R. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963.

G. B. Asheim, T. Mitra, and B. Tungodden. Sustainable recursive social welfare functions. *Economic Theory*, 49:267–292, 2012.

K. Basu and T. Mitra. Aggregating infinite utility streams with intergenerational equity: The impossibility of being paretian. *Econometrica*, 79:557–1563, 2003.

H. Bauer. *Probability Theory*. Walter de Gruyter and Co, Berlin, 1995.

A. F. Beardon and G. B. Mehta. The utility theorems of wold, debreu and arrow-hahn. *Econometrica*, 62:181–186, 1994.

A. F. Beardon, J. C. Candeal, G. Herden, E. Induráin, and G. B. Mehta. The non-existence of a utility function and the structure of non-representable preference relations. *Journal of Mathematical Economics*, 37(1):17–38, 2002.

D. Blackwell. Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2):265–272, 1953.

D. S. Bridges and G. B. Mehta. *Representations of Preference Orderings*. Springer-Verlag, Berlin, 1995.

A. Caplin and M. Dean. Rational inattention, entropy, and choice: The posterior-based approach. 2013.

H. Chernoff and S. Reiter. Selection of a distribution function to minimize an expectation subject to side conditions. Technical report, 1954.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Hoboken, NJ, 2006.

C. d'Aspremont and L. Gevers. Social welfare functionals and interpersonal comparability. In *Handbook of Social Choice and Welfare, vol 1, Elsevier,*, pages 459–541. Amsterdam, 2002.

H. de Oliveira. Axiomatic foundations for rational inattention. Unpublished manuscript, 2012.

H. de Oliveira, T. Denti, M. Mihm, and M. K. Ozbek. Rationally inattentive preferences. 2013.

G. Debreu. Representation of a preference ordering by a numerical function. In *Decision Processes, John Wiley and Sons, New*, pages 159–165. York, 1954.

E. Dekel, B. L. Lipman, and A. Rustichini. Representing preferences with a unique subjective state space. *Econometrica*, 69(4):891–934, 2001.

T. Denti. Rational inattention: A decision-theoretic perspective. Unpublished manuscript, 2012.

P. Diamond. The evaluation of infinite utility streams. *Econometrica*, 33:170–177, 1965.

D. Dillinberger and P. Sadowski. Models of subjective learning. Unpublished manuscript, 2011.

A. Ellis. Foundations for rational inattention. Unpublished manuscript, 2012.

H. Ergin and T. Sarver. A unique costly contemplation representation. *Econometrica*, 68(4):1285–1339, 2010.

P. C. Fishburn. *Utility Theory for Decision Making*. Wiley, New York, NY, 1970.

D. Garcia and J. Vanden. Information acquisition and mutual funds. *Journal of EconomicTheory*, 144:1965–95, 2005.

P. Ghirardato and M. Marinacci. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, 102, 2002.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

C. Gollier. *The Economics of Risk and Time*. MIT Press, Cambridge, MA, 2004.

M. Golosov, G. Lorenzoni, and A. Tsyvinski. Decentralized trading with private information. 2012. Unpublished manuscript.

F. Gul and W. Pesendorfer. Temptation and self-control. *Econometrica*, 69(6): 1403–1435, 2001.

C. Hara, T. Shinotsuka, K. Suzumura, and Y. Xu. On the possibility of continuous, paretian and egalitarian evaluation of infinite utility streams. *Social Choice and Welfare*, 31(2):179–191, 2008.

C. Hellwig and L. Veldkamp. Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, 76(1):223–251, 2009.

I. N. Herstein and J. Milnor. An axiomatic approach to measurable utility. *Econometrica*, 21(2):291–297, 1953.

S. Isaacson and H. Rubin. On minimizing an expectation subject to certain side conditions. Technical report, 1954.

R. E. Kihlstrom. A Bayesian exposition of Blackwell's theorem on the comparison of experiments. In M. Boyer and R. E. Kihlstrom, editors, *Bayesian Models in Economic Theory*. Elsevier Science Publishers, New York, NY, 1984.

T. C. Koopmans. Stationary ordinal utility and impatience. *Econometrica*, 28: 287–309, 1960.

D. M. Kreps. A representation theorem for "Preference for Flexibility". *Econometrica*, 47(3):565–577, 1979.

D. M. Kreps. *Notes on the Theory of Choice*. Westview Press, Boulder, Colarado, 1988.

L. Lauwers. Ordering infinite utility streams comes at the cost of a non-ramsey set. *Journal of Mathematical Economics*, 46:32–37, 2010.

J. S. Lleras. Expecting information informing expected utility: Identification of information partitions, preferences and beliefs. Unpublished manuscript, 2010.

F. Maccheroni, M. Marinacci, and A. Rustichini. Variational representation of preferences under ambiguity. *ICER working paper*, 5(2004), 2004.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

B. Mackowiak and M. Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99:769–803, 2009.

B. Mackowiak and M. Wiederholt. Business cycle dynamics under Rational Inattention. *CEPR Discussion Paper*, 7691, 2010.

Y. Masatlioglu, D. Nakajima, and E. Ozbay. Revealed attention. *American Economic Review*, 2012. forthcoming.

F. Matějka. Rationally inattentive seller: Sales and discrete pricing. 2012.

F. Matějka and A. McKay. Rational inattention to discrete choices: A new foudnation for the multinominal logit model. 2013.

T. Mitra and M. K. Ozbek. On representation and weighted utilitarian representation of preference orders on finite utility streams. Working Paper 10, CAE-05, Cornell University, 2010.

J. R. Munkres. *Topology*. Prentice Hall, Upper Saddle River, NJ, 1975.

P. Ortoleva. The price of flexibility: Towards a theory of thinking aversion. Unpublished manuscript, 2012.

B. Peleg. Utility functions for partially ordered topological spaces. *Econometrica*, 38:93–96, 1970.

L. Peng and W. Xiong. Investor attention,overconfidence and category learning. *Journal of Financial Economics*, 80(1):563–602, 2006.

T. W. Picton, D. T. Stuss, and K. C. Marshall. Attention and the brain. In S. L. Friedman, K. A. Klivington, and R. W. Peterson, editors, *The Brain, Cognition, and Education*. Academic Press, Inc, Orlando, FL, 1986.

J. S. Rustaghi. On minimizing and maximizing an integral with statistical applications. *Annals of Mathematical Statistics*, 28(2):309–328, 1957.

L. J. Savage. *The Foundations of Statistics*. John Wiley, New York, NY, 1954.

B. Simon. *Convexity: An Analytical Viewpoint*. Cambridge University Press, 2011.

H. A. Simon. Designing organizations for an information-rich world. In M. Greenberger, editor, *Computers, Communication, and the Public Interest*. The Johns Hopkins Press, Baltimore, MD, 1971.

C. A. Sims. Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49, 1998.

C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

C. A. Sims. Rational inattention: Beyond the linear-quadratic case. *The American Economic Review*, 96(2):158–163, 2006.

J. Stoye. Axioms for minimax regret choice correspondences. *Journal of Economic Theory*, 146(6):2226–2251, 2011.

L. G. Svensson. Equity among generations. *Econometrica*, 48:1251–1256, 1980.

N. Takeoka. Subjective prior over subjective states, stochastic choice, and updating. Unpublished manuscript, 2005.

S. Van Nieuwerburgh and L. Veldkamp. Information immobility and the home bias puzzle. *Journal of Finance*, 64:1187–1215, 2009.

S. Van Nieuwerburgh and L. Veldkamp. Information acquisition and under-diversification. *Review of Economic Studies*, 77:779–805, 2010.

L. Veldkamp. *Information Choice in Macroeconomics and Finance*. Princeton University Press, Princeton, NJ, 2011.

L. Veldkamp and J. Wolfers. Aggregate shocks or aggregate information? costly information and business cycle comovement. *Journal of Monetary Economics*, 54:37–55, 2007.

M. Voorneveld and J. W. Weibull. Outer measure and utility. Working paper, Series in Economics and Finance 704, Stockholm School of Economics, 2009.

J. W. Weibull. Discounted-value representations of temporal preferences. *Mathematics of Operations Research*, 10:244–250, 1985.

M. Wiederholt. Rational inattention. In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2010.

H. Wold. A synthesis of pure demand analysis, i, ii and iii. *Scandinavisk Aktuarietidskrift*, 26(85-118):69–120, 1943.

M. Woodford. Information-constrained state-dependent pricing. *Journal of Monetary Economics*, 56, Supplement:S100–S124, 2009.

W. R. Zame. Can intergenerational equity be operationalized? *Theoretical Economics*, 2:187–202, 2007.